TEST DE HIPÓTESIS

1 Introducción

En este tema vamos a estudiar el concepto de test de hipótesis así como la intuición subyacente a algunos test de hipótesis clásicos. Una vez vistos los conceptos básicos presentaremos un resultado teórico (el lema de Neyman–Pearson) que nos va a permitir obtener los "mejores" tests de manera analítica y exacta. Como nuestro objetivo en la asignatura no es llevar a cabo un estudio teórico de los test de hipótesis ni tampoco obtenerlos con un alto rigor matemático, la presentación de la mayoría de los tests de hipótesis se llevará a cabo desde una perspectiva intuitiva.

A lo largo del tema iremos obteniendo tests de hipótesis para proporciones (parámetro p de una distribución de Bernouilli), para medias aritméticas (parámetro μ de una distribución normal), para chequear la independencia entre dos distribuciones (basándonos en la distribución chi–cuadrado), para chequear el ajuste de una distribución de probabilidad a un conjunto de datos muestrales (basándonos en la distribución chi–cuadrado o en el estadístico de Kolmogoroff–Smirnov) y finalmente algunos tests que no asumen una distribución de probabilidad concreta sobre los datos.

2 Concepto de test de hipótesis

El concepto de *test de hipótesis* es básico y fundamental en la demostración de conjeturas dentro de las denominadas ciencias experimentales. Podemos decir que el razonamiento científico en dichas ciencias experimentales descansa sobre la idea subyacente a los test de hipótesis. De hecho dicho razonamiento científico utilizará la demostración matemática para la construcción de una teoría científica en las denominadas ciencias exactas, mientras que en las ciencias experimentales nos vemos abocados a recoger datos –provenientes de experimentos– los cuales una vez analizados permitirán aceptar o rechazar una determinada hipótesis que tengamos acerca del dominio motivo de estudio.

Al igual que en los dos temas anteriores –estimación puntual y estimación por intervalos– en este tema partiremos de una muestra aleatoria simple extraida de la población, con la cual pretendemos "decir algo" acerca de la población. Es decir volvemos en este tema a efectuar un ejercicio de *inferencia estadística*, si bien el alcance de los resultados relacionados con este tema es mucho mayor que lo obtenido en temas anteriores, ya que a partir de los resultados obtenidos en los tests de hipótesis vamos a poder decidir acerca de la aceptación o el rechazo de distintas hipótesis.

Conviene aclarar que dicha decisión va a estar sometida a errores. Es decir puede ocurrir que basándonos en la información recogida en la muestra decidamos aceptar una determinada hipótesis, la cual en realidad es falsa, o bien puede darse tambien el caso contrario en el que teniendo en cuenta los resultados de la muestra decidamos rechazar la hipótesis cuando en realidad dicha hipótesis es verdadera. Obviamente pueden darse otros dos casos, como son que basándonos en la muestra aceptemos la hipótesis siendo en realidad dicha hipótesis correcta, o que alternativamente basándonos en la muestra rechazemos la hipótesis siendo la misma falsa. La tabla 3.1 recoge las 4 posibilidades analizadas.

Tabla 3.1: Las cuatro posibilidades al efectuar un test de hipótesis

		H_0 Verdadera	H_0 Falsa
Decisión en base a la muestra	Aceptar	Correcta	Error
	Rechazar	Error	Correcta

Aunque no lo hemos dicho hasta el momento, cuando testamos una determinada hipótesis, en realidad la testamos o contrastamos frente a otra hipótesis que se acostumbra a llamar *hipótesis alternativa*. La terminología que se usa para la primera hipótesis es *hipótesis nula*. Es decir que

observada una muestra nos tenemos que decidir sobre la veracidad de la hipótesis nula al contrastar dicha hipótesis frente a la denominada hipótesis alternativa.

Llegados a este punto de la presentación conviene aclarar que ambas hipótesis –nula y alternativa– no están actuando de manera simétrica. Con esto se quiere expresar que la información proveniente de la muestra va a ser utilizada en aras de decidir si con la misma tenemos la evidencia suficiente como para rechazar la hipótesis nula. Entiéndase que el rechazo de la hipótesis nula no es equivalente a la aceptación de la hipótesis alternativa. Para aceptar esta última deberemos de ponerla a testar como hipótesis nula.

Para entender de manera intuitiva el contenido del párrafo anterior recurriremos a la analogía existente entre un test de hipótesis y un juicio. Para construir la analogía pensemos que el juez debe de decidir si el acusado es inocente o culpable basándose en la información que le presentan tanto el abogado defensor como el fiscal. Siguiendo con el simil, la inocencia del acusado actúa como hipótesis nula, la cual se constrasta frente a la hipótesis alternativa, culpabilidad en el juicio. La falta de simetria entre ambas hipótesis es evidente en el ejemplo del juicio, ya que todos sabemos que "mientras no se demuestre lo contrario el acusado es inocente". Dicho de otra manera, mientras que la evidencia que presente el fiscal no sea suficiente para probar –en criterio del juez– la culpabilidad del acusado, este será calificado como inocente.

El ejemplo sirve tambien para ilustrar las cuatro posibilidades que se muestran en la tabla 3.1, y que en este caso consisten en:

- declarar inocente a un individuo que en realidad es inocente
- declarar inocente a un individuo que en realidad es culpable
- declarar culpable a un individuo que en realidad es inocente
- declarar culpable a un individuo que en realidad es culpable.

Con objeto de ir introduciendo la terminología y la notación matemática relacionada con este tema, vamos a considerar el ejemplo que se introdujo en el tema 1 acerca del algoritmo heurístico de optimización en el que medíamos el tiempo de ejecución así como el éxito o fracaso del mismo, en alusión a que con dicho algoritmo se alcanzase el óptimo global o no.

Definición 3.1
Dada una variable aleatoria X, con ley de probabilidad conocida, ($P(X = x; \theta)$ en caso de que X sea discreta, o $f_X(x; \theta)$ en caso de que X sea continua) dependiente de un parámetro desconocido θ. Extraemos una muestra aleatoria simple, de tamaño n, $x_1, \ldots, x_n$ de X, la cual puede ser vista como la realización de n variables aleatorias independientes e idénticamente distribuidas, $X_1, \ldots, X_n$. A partir de dicha muestra pretendemos contrastar la hipótesis de que el valor de θ es θ_0 ($H_0 : \theta = \theta_0$, *hipótesis nula*) frente a la hipótesis de que el valor de θ es θ_1 ($H_1 : \theta = \theta_1$, *hipótesis alternativa*), lo cual se expresará de la siguiente manera:

$$\begin{cases} H_0 : \theta = \theta_0 & \text{hipótesis nula,} \\ H_1 : \theta = \theta_1 & \text{hipótesis alternativa.} \end{cases} \tag{1}$$

Ejemplo 3.1
Para el caso del heurístico estocático de optimización nos podemos plantear por el siguiente test de hipótesis en relación con la variable aleatoria midiendo el tiempo de ejecución, la cual supongamos que sigue un modelo $\mathcal{N}(\mu, 100)$.

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 50 \end{cases} \tag{2}$$

De manera análoga para la variable aleatoria midiendo el éxito o fracaso del algoritmo, la cual sigue un modelo $B(1, p)$, podemos plantear el contraste:

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p = 0.9 \end{cases} \tag{3}$$

Definición 3.2
Una hipótesis se dice que es una *hipótesis simple* si a partir de ella la distribución de probabilidad resulta totalmente especificada. En caso contrario diremos que se trata de una *hipótesis compuesta*.

Ejemplo 3.2
(a) Para el caso de la $\mathcal{N}(\mu, 100)$ anterior:
$H : \mu = 60$ es una hipótesis simple
$H : \mu < 60$ es una hipótesis compuesta

(b) Para el caso de la $B(1, p)$ anterior:
$H : p = 0.5$ es una hipótesis simple
$H : p > 0.8$ es una hipótesis compuesta

(c) Para el caso de la $\mathcal{N}(\mu, \sigma)$ con σ desconocida:
$H : \mu = 60$ es una hipótesis compuesta
$H : \mu < 60$ es una hipótesis compuesta

Definición 3.3
Dado un test de hipótesis, la *región de aceptación* de dicho test está constituida por un conjunto de muestras a partir de las cuales se decide aceptar la hipótesis nula. De manera análoga, la *región de rechazo* correspondiente a un test está constituida por las muestras a partir de las cuales se decide rechazar la hipótesis nula frente a la hipótesis alternativa.

Denotaremos la región de aceptación por $R.A.$ La región de rechazo se denotará por $R.R.$ Según las definiciones anteriores, tenemos que:

$(x_1, \ldots, x_n) \in R.A. \Leftrightarrow H_0$ se acepta al contrastarla frente a H_1

$(x_1, \ldots, x_n) \in R.R. \Leftrightarrow H_0$ se rechaza al contrastarla frente a H_1

Ejemplo 3.3
(a) Para el test de hipótesis:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 50 \end{cases} \tag{4}$$

introducido en el ejemplo 3.1, una región de aceptación intuitiva puede estar constituida por aquellas muestras en las que la media aritmética sea mayor que 55. Es decir:

$$R.A. = \{(x_1, \ldots, x_n) \mid \overline{X} > 55\} \tag{5}$$

o alternativamente:

$$R.R. = \{(x_1, \ldots, x_n) \mid \overline{X} \leq 55\} \tag{6}$$

(b) Para el test de hipótesis:

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p = 0.9 \end{cases} \tag{7}$$

planteado en el ejemplo 3.1, una región de aceptación intuitiva puede estar constituida por aquellas muestras en las que la frecuencia relativa de éxito, $\hat{p}$, ha sido menor que 0.7. Es decir:

$$R.A. = \{(x_1, \ldots, x_n) \mid \hat{p} < 0.7\} \tag{8}$$

o alternativamente:

$$R.R. = \{(x_1, \ldots, x_n) \mid \hat{p} \geq 0.7\} \tag{9}$$

Tal y como vemos en el ejemplo 3.3, la región de rechazo resulta ser el complementario de la región de aceptación. Por otra parte para establecer cualquiera de ellas necesitamos conocer la ley de probabilidad de un *estadístico* ($\overline{X}$ en el ejemplo 3.3(a), y $\hat{p}$ en el ejemplo 3.3(b)) a partir del cual se construye el contraste.

Se ha comentado anteriormente que cualquier decisión tomada bajo incertidumbre (en el caso de los test de hipótesis la decisión se basa en la información recogida en la muestra) tiene asociada dos tipos de error. Dichos errores los hemos introducido de manera intuitiva en la tabla 3.1. Veamos a continuación una definición precisa de los mismos.

Definición 3.4
Dado un test de hipótesis, llamaremos *error de tipo I* al que se comete al rechazar la hipótesis nula cuando esta es en realidad verdadera. De manera análoga definimos el *error de tipo II* como el cometido al aceptar la hipótesis nula cuando esta es en realidad falsa.

En el caso en que tanto la hipótesis nula como la alternativa sean simples, podemos cuantificar la probabilidad de cometer cualquiera de los dos errores anteriores. Denotaremos por α a la probabilidad de cometer un error de tipo I, mientras que β denotará la probabilidad de cometer un error de tipo II. Es decir:

$$\alpha = P_{H_0}[(x_1, \ldots, x_n) \in R.R.] \tag{10}$$

$$\beta = P_{H_1}[(x_1, \ldots, x_n) \in R.A.] \tag{11}$$

Se intuye que las regiones de aceptación (o equivalentemente las regiones de rechazo) deberán de construirse de manera que α y β sean lo menor posible.

Ejemplo 3.4
Supongamos que $X \to \mathcal{N}(\mu, 100)$ y a partir de una muestra aleatoria de tamaño $n = 49$ queremos efectuar el siguiente test de hipótesis:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 50 \end{cases} \tag{12}$$

(a) Vamos a partir de la región de aceptación $R.A. = \{(x_1, \ldots, x_n) \mid \overline{X} > 55\}$, y calcularemos las probabilidades de cometer un error de tipo I y un error de tipo II, es decir α y β respectivamente.

Bajo H_0, es decir asumiendo que H_0 es cierta, se tiene que:

$$\overline{X} \overset{H_0}{\to} \mathcal{N}(60, \frac{100}{7}) \tag{13}$$

mientras que bajo H_1, es decir ahora suponiendo que H_1 es cierta, obtenemos que:

$$\overline{X} \overset{H_1}{\to} \mathcal{N}(50, \frac{100}{7}) \tag{14}$$

De ahí que:

$$\alpha = P_{H_0}[(x_1, \ldots, x_{49}) \in R.R.] = P_{H_0}[\overline{X} \leq 55] = P[\mathcal{N}(60, \frac{100}{7}) \leq 55] =$$

$$P[\mathcal{N}(0,1) \leq \frac{55 - 60}{\frac{100}{7}}] = P[\mathcal{N}(0,1) \leq -0.35] = P[\mathcal{N}(0,1) > 0.35] = 1 - 0.6368 = 0.3632. \tag{15}$$

Análogamente:

$$\beta = P_{H_1}[(x_1, \ldots, x_{49}) \in R.A.] = P_{H_1}[\overline{X} > 55] = P[\mathcal{N}(50, \frac{100}{7}) > 55] =$$

$$P[\mathcal{N}(0,1) \leq \frac{55-50}{\frac{100}{7}}] = P[\mathcal{N}(0,1) > 0.35] = 0.3632. \quad (16)$$

La interpretación de los valores de α y β obtenidos nos indica que aproximadamente el 36 por ciento de las veces en las que rechazamos la hipótesis nula –al salirnos una media muestral inferior a 55– nos estamos equivocando, de la misma forma que tambien nos equivocamos en aproximadamente el 36 por ciento de las veces en las que aceptamos la hipótesis alternativa –al salirnos una media muestral superior a 55–.

(b) Procedemos ahora en sentido inverso. Es decir fijamos el valor de α (por ejemplo $\alpha = 0.05$) y a partir de dicho valor obtenemos la región de rechazo y el valor de β.

Teniendo en cuenta que la región de rechazo intuitiva es de la forma:

$$R.R. = \{(x_1, \ldots, x_{49}) \mid \overline{X} \leq K\} \quad (17)$$

tratemos de determinarla para $\alpha = 0.05$. Tenemos por tanto que:

$$0.05 = P_{H_0}[(x_1, \ldots, x_{49}) \in R.R.] = P_{H_0}[\overline{X} \leq K] = P[\mathcal{N}(60, \frac{100}{7}) \leq K] = P[\mathcal{N}(0,1) \leq \frac{K-60}{\frac{100}{7}}] \quad (18)$$

Consultando las tablas de la distribución normal, se tiene que $\frac{K-60}{\frac{100}{7}} = -1.64$, de ahí que $K = 46.72$.

La interpretación del resultado obtenido es como sigue: si partiendo de una muestra de tamaño 49, extraida de una población normal con desviación típica 100, tratamos de contrastar la hipótesis nula $H_0 : \mu = 60$, frente a la hipótesis alternativa $H_1 : \mu = 50$, a partir de la decisión consistente en rechazar la hipótesis nula siempre y cuando la media aritmética sea inferior a 46.72, una vez de cada veinte estamos cometiendo el error de rechazar dicha hipótesis nula cuando en realidad es verdadera.

Calculemos el β asociado:

$$\beta = P_{H_1}[(x_1, \ldots, x_{49}) \in R.A.] = P_{H_1}[\overline{X} > 46.72] = P[\mathcal{N}(50, \frac{100}{7}) > 46.72] =$$

$$P[\mathcal{N}(0,1) > \frac{46.72-50}{\frac{100}{7}}] = P[\mathcal{N}(0,1) > -2.286] = P[\mathcal{N}(0,1) < 2.286] = 0.98899. \quad (19)$$

La interpretación semántica de esta probabilidad es análoga a la hecha para α.

Como vemos con este ejemplo, al tratar de que la probabilidad de cometer un error de tipo I se reduzca, lo que se obtiene es que la probabilidad de cometer un error de tipo II se incrementa.

De todas formas en este ejemplo las probabilidades de cometer tanto un error de tipo I como un error de tipo II son muy grandes debido fundamentalmente al hecho de que la desviación típica de la distribución inicial es muy alta. Vamos a plantearnos la repetición del ejemplo anterior, ahora con una distribución normal con desviación típica igual a 20.

Ejemplo 3.5
Sea $X \to \mathcal{N}(\mu, 20)$. A partir de una muestra aleatoria de tamaño $n = 49$ vamos a calcular α y β para el siguiente test de hipótesis:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 50 \end{cases} \tag{20}$$

(a) Partiremos de una región de aceptación intuitiva utilizando el estadístico $\overline{X}$. Sea $R.A. = \{(x_1, \ldots, x_{49}) \mid \overline{X} > 55\}$.

Ahora tenemos que las distribuciones de probabilidad del estadístico $\overline{X}$, bajo H_0 y bajo H_1 son respectivamente:

$$\overline{X} \overset{H_0}{\rightarrow} \mathcal{N}(60, \frac{20}{7}) \tag{21}$$

$$\overline{X} \overset{H_1}{\rightarrow} \mathcal{N}(50, \frac{20}{7}) \tag{22}$$

Por tanto:

$$\alpha = P_{H_0}[(x_1, \ldots, x_{49}) \in R.R.] = P_{H_0}[\overline{X} \leq 55] = P[\mathcal{N}(60, \frac{20}{7}) \leq 55] =$$

$$P[\mathcal{N}(0,1) \leq \frac{55-60}{\frac{20}{7}}] = P[\mathcal{N}(0,1) \leq -1.75] = P[\mathcal{N}(0,1) > 1.75] = 1 - 0.9599 = 0.0401. \tag{23}$$

Análogamente calculamos:

$$\beta = P_{H_1}[(x_1, \ldots, x_{49}) \in R.A.] = P_{H_1}[\overline{X} > 55] = P[\mathcal{N}(50, \frac{20}{7}) > 55] =$$

$$P[\mathcal{N}(0,1) > \frac{55-50}{\frac{20}{7}}] = P[\mathcal{N}(0,1) > 1.75] = 0.0401. \tag{24}$$

Vemos por tanto que una reducción en la variabilidad de la distribución asumida trae como consecuencia que los valores de α y β se reduzcan.

(b) Al igual que nos hemos planteado en el ejemplo anterior tratemos ahora de obtener la región de aceptación correspondiente a un valor concreto de α, para posteriormente calcular el β asociado a dicha región de aceptación.

Si fijamos el valor de α en 0.01, obtenemos:

$$0.01 = \alpha = P_{H_0}[(x_1, \ldots, x_{49}) \in R.R.] = P_{H_0}[\overline{X} \leq K] =$$

$$P[\mathcal{N}(60, \frac{20}{7}) \leq K] = P[\mathcal{N}(0,1) \leq \frac{K-60}{\frac{20}{7}}] \tag{25}$$

De las tablas de la $\mathcal{N}(0,1)$ se obtiene que $\frac{K-60}{\frac{20}{7}} = -2.33$.

De ahí que $K = -2.33\frac{20}{7} + 60 = 53.35$.

Calculemos el β asociado a la región de aceptación correspondiente:

$$\beta = P_{H_1}[(x_1, \ldots, x_{49}) \in R.A.] = P_{H_1}[\overline{X} > 53.35] = P[\mathcal{N}(50, \frac{20}{7}) > 53.35] =$$

$$P[\mathcal{N}(0,1) > \frac{53.355-50}{\frac{20}{7}}] = P[\mathcal{N}(0,1) > 1.17] = 0.1210. \tag{26}$$

Volvemos a obtener un resultado análogo al que hemos conseguido en el ejemplo anterior. Es decir al tratar de reducir el α la consecuencia es que aumenta el β.

La figura 3.1 ilustra tal comportamiento.

Figura 3.1: Visualización del hecho de que no es posible minimizar a la vez α y β.

Figura 3.1 Veamos en el ejemplo que sigue la manera de calcular α y β en el caso de una distribución discreta.

Ejemplo 3.6
Consideremos el test de hipótesis introducido en el ejemplo 3.1, para el parámetro p de una $B(1,p)$:

$$\begin{cases} H_0 : p = 0.5 \\ H_1 : p = 0.9 \end{cases} \tag{27}$$

A partir de una muestra de tamaño $n = 10$ una región de rechazo intuitiva es:

$$R.R. = \{(x_1, \ldots, x_{10}) \mid \sum_{i=1}^{10} X_i > 7\} \tag{28}$$

Es decir rechazo la hipótesis nula ($H_0 : p = 0.5$) frente a la alternativa ($H_1 : p = 0.9$) cuando en 10 experimentos obtengo 8, 9 o 10 éxitos.

Calculemos α y β para dicha región de rechazo. Para ello tenemos en cuenta que conocemos la distribución del estadístico $\sum_{i=1}^{10} X_i$ a partir del cual se ha construido el test: En concreto, se tiene:

$$\sum_{i=1}^{10} X_i \overset{H_0}{\to} B(10, 0.5) \tag{29}$$

$$\sum_{i=1}^{10} X_i \overset{H_1}{\to} B(10, 0.9) \tag{30}$$

La probabilidad de cometer un error de tipo I se calcula de la siguiente manera:

$$\alpha = P_{H_0}[(x_1, \ldots, x_{10}) \in R.R.] = P_{H_0}[\sum_{i=1}^{10} X_i > 7] = P[B(10, 0.5) > 7] = \sum_{i=8}^{10} \binom{10}{i} (0.5)^i (0.5)^{10-i} \tag{31}$$

De igual manera calculamos:

$$\beta = P_{H_1}[(x_1, \ldots, x_{10}) \in R.A.] = P_{H_1}[\sum_{i=1}^{10} X_i \leq 7] = P[B(10, 0.9) \leq 7] = \sum_{i=0}^{7} \binom{10}{i} (0.9)^i (0.1)^{10-i} \tag{32}$$

En la tabla 3.2 volvemos a resumir –al igual que hemos hecho en la tabla 3.1– las cuatro posibilidades existentes al efectuar un test de hipótesis, indicando ahora los conceptos de región de aceptación, región de rechazo, error de tipo I y error de tipo II.

Tabla 3.2: Esquema relacionado con un test de hipótesis.

		H_0 Verdadera	H_0 Falsa
Decisón en base a la muestra	Aceptar $(x_1, \ldots x_n) \in R.A.$	Correcta	Error de tipo II $\beta = P_{H_1}[(x_1, \ldots x_n) \in R.A.]$
	Rechazar $(x_1, \ldots x_n) \in R.R.$	Error de tipo I $\alpha = P_{H_0}[(x_1, \ldots x_n) \in R.R.]$	Correcta

Conviene matizar que el cálculo de α tan sólo es posible para el caso en que la hipótesis nula sea una hipótesis simple. De manera análoga para calcular β necesitamos que la hipótesis alternativa sea simple.

3 Lema de Neyman–Pearson

Hemos comentado en el apartado anterior que una buena región de aceptación debería de tener asociados valores de α y β pequeños, pero por otra parte en los ejemplos 3.4 y 3.5 se ha visto que al tratar de reducir α la consecuencia es un incremento de β. Por este motivo nos vemos obligados a definir el concepto de *región de rechazo óptima a nivel* α.

Definición 3.5
Sea $R.R.^*$ una región de rechazo para un determinado test de hipótesis, con hipótesis nula e hipótesis alternativa simples. Denotamos por α^* y β^* respectivamente a las probabilidades de cometer un error de tipo I y tipo II a partir de $R.R.^*$. Se dice que $R.R.^*$ es una *región de rechazo óptima a nivel* α si se verifica que para cualquier otra región de rechazo $R.R.$ con α y β asociados, tales que $\alpha = \alpha^*$ se tiene que $\beta^* \leq \beta$.

Es decir la región de rechazo óptima a nivel α es aquella que tiene menor probabilidad de cometer un error de tipo II asociado entre todas las regiones de rechazo con el mismo nivel α.

A continuación vamos a enunciar un resultado a partir del cual se va a poder obtener la región de rechazo óptima a nivel α, para tests de hipótesis con hipótesis nula e hipótesis alternativa simples.

Teorema 3.1 (Lema de Neyman-Pearson)
Sea X una variable aleatoria con distribución de probabilidad conocida, dependiente de un parámetro θ desconocido. Denotamos por $L(x_1, \ldots, x_n; \theta)$ a la función de verosimilitud asociada a la muestra de tamaño n, $x_1, \ldots, x_n$ correspondiente a una realización de las variables aleatorias $X_1, \ldots, X_n$ las cuales son independientes y están idénticamente distribuidas que X.

La región de rechazo óptima a nivel α, $R.R.^*$ para el test de hipótesis:

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases} \tag{33}$$

es de la forma siguiente:

$$R.R.^* = \{(x_1, \ldots, x_n) \mid \frac{L(x_1, \ldots, x_n; \theta_0)}{L(x_1, \ldots, x_n; \theta_1)} \leq K\} \tag{34}$$

siendo K una constante a calcular para que la región de rechazo sea de nivel α.

El resultado del lema de Neyman-Pearson es intuitivo ya que viene a decir que la evidencia proveniente de una muestra va a servirnos para rechazar una determinada hipótesis nula si la verosimilitud que la muestra asigna a dicha hipótesis nula es menor que el producto entre una constante y la verosimilitud asignada por la muestra a la hipótesis alternativa.

Veamos dos ejemplos de aplicación del lema de Neyman-Pearson a las distribuciones normal y de Bernouilli respectivamente, constatando en ambos casos que las regiones de rechazo óptimas obtenidas como aplicación del lema anterior coinciden con las que podemos plantear de manera intuitiva.

Ejemplo 3.7
Sea X una variable aleatoria normal con parámetros μ y σ, siendo este último conocido. A partir de una muestra aleatoria simple de tamaño n extraida de X vamos a construir la región de rechazo óptima a nivel α para el test de hipótesis:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases} \tag{35}$$

Aplicaremos para ello el lema de Neyman-Pearson y tendremos en cuenta dos posibilidades en relación con μ_0 y μ_1, es decir tanto el caso en que $\mu_0 < \mu_1$ como el caso en que $\mu_0 > \mu_1$.

Según el lema anterior:

$$(x_1, \ldots, x_n) \in R.R.^* \Leftrightarrow \frac{L(x_1, \ldots, x_n; \mu_0)}{L(x_1, \ldots, x_n; \mu_1)} \leq K \Leftrightarrow$$

$$\frac{(\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2}\sum_{i=1}^n (\frac{x_i-\mu_0}{\sigma})^2}}{(\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2}\sum_{i=1}^n (\frac{x_i-\mu_1}{\sigma})^2}} \leq K \Leftrightarrow$$

$$e^{\frac{1}{2\sigma^2}(\sum_{i=1}^n (x_i-\mu_1)^2 - \sum_{i=1}^n (x_i-\mu_0)^2)} \leq K \Leftrightarrow$$

$$\frac{1}{2\sigma^2}(\sum_{i=1}^n (x_i-\mu_1)^2 - \sum_{i=1}^n (x_i-\mu_0)^2) \leq \ln K \Leftrightarrow \sum_{i=1}^n (x_i^2 - 2x_i\mu_1 + \mu_1^2) - \sum_{i=1}^n (x_i^2 - 2x_i\mu_0 + \mu_0^2) \leq 2\sigma^2 \ln K \Leftrightarrow$$

$$\sum_{i=1}^n x_i^2 - 2\mu_1 \sum_{i=1}^n x_i + n\mu_1^2 - \sum_{i=1}^n x_i^2 + 2\mu_0 \sum_{i=1}^n x_i - n\mu_0^2 \leq 2\sigma^2 \ln K \Leftrightarrow$$

$$2(\mu_0 - \mu_1) \sum_{i=1}^n x_i + n(\mu_1^2 - \mu_0^2) \leq 2\sigma^2 \ln K \Leftrightarrow$$

$$(\mu_0 - \mu_1)\sum_{i=1}^{n} x_i \leq \frac{2\sigma^2 \ln K - n(\mu_1^2 - \mu_0^2)}{2}$$

Denotando por K^* a $\frac{2\sigma^2 \ln K - n(\mu_1^2 - \mu_0^2)}{2}$ hemos obtenido que:

$$(x_1, \ldots, x_n) \in RR^* \Leftrightarrow (\mu_0 - \mu_1)\sum_{i=1}^{n} x_i \leq K^* \tag{36}$$

Consideremos a continuación los dos casos anteriormente comentados:

(a) Si $\mu_0 > \mu_1$ obtenemos:

$$(x_1, \ldots, x_n) \in RR^* \Leftrightarrow \sum_{i=1}^{n} x_i \leq \frac{K^*}{(\mu_0 - \mu_1)} \tag{37}$$

o equivalentemente:

$$RR^* = \{(x_1, \ldots, x_n) \mid \overline{X} \leq c\} \tag{38}$$

Nótese que c debe calcularse para que la región de rechazo tenga el nivel α previamente prefijado. Obsérvese tambien que esta región de rechazo es de la forma de las regiones de rechazo expresadas en los ejemplos 3.4 y 3.5.

(b) Si $\mu_0 < \mu_1$ obtenemos:

$$(x_1, \ldots, x_n) \in RR^* \Leftrightarrow \sum_{i=1}^{n} x_i \geq \frac{K^*}{(\mu_0 - \mu_1)} \tag{39}$$

o equivalentemente:

$$RR^* = \{(x_1, \ldots, x_n) \mid \overline{X} \geq c\} \tag{40}$$

Tambien en este caso se debe de calcular c para que la región de rechazo tenga nivel α.

Ejemplo 3.8

Sea X una variable aleatoria de Bernouilli de parámetro p desconocido. A partir de una muestra aleatoria simple de tamaño n extraida de X vamos a construir la región de rechazo óptima a nivel α para el test de hipótesis:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 \end{cases} \tag{41}$$

Al igual que en el ejemplo anterior consideraremos dos casos según que p_0 sea mayor o menor que p_1.

Aplicando el lema de Neyman-Pearson, obtenemos:

$$(x_1, \ldots, x_n) \in R.R.^* \Leftrightarrow \frac{L(x_1, \ldots, x_n; p_0)}{L(x_1, \ldots, x_n; p_1)} \leq K \Leftrightarrow$$

$$\frac{\prod_{i=1}^{n} p_0^{x_i}(1-p_0)^{1-x_i}}{\prod_{i=1}^{n} p_1^{x_i}(1-p_1)^{1-x_i}} \leq K \Leftrightarrow \frac{p_0^{\sum_{i=1}^{n} x_i}(1-p_0)^{n-\sum_{i=1}^{n} x_i}}{p_1^{\sum_{i=1}^{n} x_i}(1-p_1)^{n-\sum_{i=1}^{n} x_i}} \leq K \Leftrightarrow$$

$$(\frac{p_0}{p_1})^{\sum_{i=1}^{n} x_i}(\frac{1-p_0}{1-p_1})^{n-\sum_{i=1}^{n} x_i} \leq K \Leftrightarrow (\frac{p_0}{p_1})^{\sum_{i=1}^{n} x_i}(\frac{1-p_0}{1-p_1})^{n}(\frac{1-p_1}{1-p_0})^{\sum_{i=1}^{n} x_i} \leq K \Leftrightarrow$$

$$(\frac{p_0(1-p_1)}{p_1(1-p_0)})^{\sum_{i=1}^n x_i} \leq K(\frac{1-p_1}{1-p_0})^n \Leftrightarrow (\sum_{i=1}^n x_i)\ln(\frac{p_0(1-p_1)}{p_1(1-p_0)}) \leq \ln K + n\ln(\frac{(1-p_1)}{(1-p_0)})$$

Denotaremos por $K^* = \ln K + n\ln(\frac{(1-p_1)}{(1-p_0)})$. Estudiando a partir de ahora cada uno de los dos casos planteados por separado se tiene:

(a) Si $p_0 > p_1$ obtenemos que $1-p_1 > 1-p_0$ y por tanto $p_0(1-p_1) > p_1(1-p_0)$. Es decir $\frac{p_0(1-p_1)}{p_1(1-p_0)} > 1$ y de ahí que: $\ln\frac{p_0(1-p_1)}{p_1(1-p_0)} > 0$. Por tanto en este caso la región de rechazo óptima, a nivel α, $R.R.^*$, será:

$$(x_1,\ldots,x_n) \in RR^* \Leftrightarrow \sum_{i=1}^n x_i \leq \frac{K^*}{\ln(\frac{p_0(1-p_1)}{p_1(1-p_0)})} \tag{42}$$

o equivalentemente:

$$RR^* = \{(x_1,\ldots,x_n) \mid \sum_{i=1}^n x_i \leq c\} \tag{43}$$

Al igual que en el ejemplo anterior, c se determinará para que la región de rechazo tenga una probabilidad de cometer un error de tipo I de α.

(b) Si $p_0 < p_1$, repitiendo el razonamiento anterior se obtiene:

$$RR^* = \{(x_1,\ldots,x_n) \mid \sum_{i=1}^n x_i \geq c\} \tag{44}$$

4 Test de la razón de verosimilitud

Enunciamos a continuación el denominado *test de la razón de verosimilitud* el cual generaliza el lema de Neyman–Pearson a situaciones en las que bien la hipótesis nula o la hipótesis alternativa son compuestas. El motivo por el que se presenta el resultado del test de la razón de la verosimilitud es debido a que buena parte de los test de hipótesis con los que trabajaremos en este tema se han obtenido como aplicación de dicho test.

Teorema 3.2
Sea X una variable aleatoria con distribución de probabilidad conocida, dependiente de un parámetro $\boldsymbol{\theta}$ desconocido. Denotamos por $L(x_1,\ldots,x_n;\boldsymbol{\theta})$ a la función de verosimilitud asociada a la muestra de tamaño n, $x_1,\ldots,x_n$ correspondiente a una realización de las variables aleatorias $X_1,\ldots,X_n$ las cuales son independientes y están idénticamente distribuidas que X.

La región de rechazo $R.R.$ para el test de hipótesis:

$$\begin{cases} H_0 : \boldsymbol{\theta} \in \Omega_1 \\ H_1 : \boldsymbol{\theta} \in \Omega \backslash \Omega_1 \end{cases} \tag{45}$$

es de la forma siguiente:

$$R.R.^* = \{(x_1,\ldots,x_n) \mid \frac{\max_{\boldsymbol{\theta}\in\Omega_1} L(x_1,\ldots,x_n;\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Omega} L(x_1,\ldots,x_n;\boldsymbol{\theta})} \leq K\} \tag{46}$$

Según el tipo de problema puede resultar más o menos dificultoso encontrar la distribución de probabilidad –tanto bajo Ω_1, como bajo $\Omega\backslash\Omega_1$– del estadístico utilizado para construir la región de rechazo. En todo caso se conoce que la distribución asintótica (cuando n aumenta) del cociente entre los máximos de la función de verosimilitud tiende hacia una ley de probabilidad chi–cuadrado.

Más concretamente:

$$-2\ln \frac{\max_{\boldsymbol{\theta}\in\Omega_1} L(x_1,\ldots,x_n;\boldsymbol{\theta})}{\max_{\boldsymbol{\theta}\in\Omega} L(x_1,\ldots,x_n;\boldsymbol{\theta})} \overset{n\to\infty}{\longrightarrow} \chi^2_{k-r} \tag{47}$$

siendo $k = \dim\Omega$ y $r = \dim\Omega_1$.

5 Test de hipótesis para la esperanza matemática de una distribución normal

En este apartado se presentarán enunciados como teoremas diversos resultados relacionados con diferentes tests de hipótesis establecidos para la esperanza matemática de una distribución normal. Los resultados se obtienen al aplicar el test de la razón de verosimilitud.

5.1 Varianza conocida

Teorema 3.3
Sea X una variable aleatoria normal con parámetros μ y σ, siendo este último conocido. A partir de una muestra aleatoria simple de tamaño n extraida de X la región de rechazo obtenida al aplicar el test de la razón de verosimilitud al test de hipótesis:

1.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 > \mu_o) \end{cases} \tag{48}$$

es: $R.R. = \{(x_1,\ldots,x_n) \mid \overline{X} > c\}$

2.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 < \mu_o) \end{cases} \tag{49}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \overline{X} < c\}$

3.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 \neq \mu_o) \end{cases} \tag{50}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \overline{X} < c_1, \overline{X} > c_2\}$ con $c_1 < c_2$.

Obsérvese que en los dos primeros casos la región de rechazo es hacia un lado, mientras que en el tercer caso es hacia dos lados. Esto nos permite introducir la terminología de *test de un lado* o *test de dos lados* en función de la forma de la región de rechazo asociada al test.

Es importante darse cuenta de otra diferencia fundamental entre las regiones de rechazo para los dos primeros casos al compararlos con la región de rechazo para el tercer caso. Mientras que en los dos primeros conocemos la ley de distribución del estadístico $\overline{X}$, a partir del cual se construyó la región de rechazo, tanto para H_0 como para H_1, en el tercer caso al ser la hipótesis alternativa compuesta, el estadístico $\overline{X}$ sigue una ley de probabilidad que no está determinada, sino que depende del punto de H_1 que se considere.

Podemos tambien plantearnos la obtención intuitiva de las regiones de rechazo correspondientes por ejemplo a los siguientes tets de hipótesis:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \tag{51}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \tag{52}$$

5.2 Varianza desconocida

Teorema 3.4
Sea X una variable aleatoria normal con parámetros μ y σ, siendo este último desconocido. A partir de una muestra aleatoria simple de tamaño n extraida de X la región de rechazo obtenida al aplicar el test de la razón de verosimilitud al test de hipótesis:

1.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 > \mu_0) \end{cases} \tag{53}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \frac{\overline{X} - \mu_0}{S_{n-1}/\sqrt{n}} > c\}$ con c verificando que: $P(t_{n-1} > c) = \alpha$.

2.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 > \mu_0) \end{cases} \tag{54}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \frac{\overline{X} - \mu_0}{S_{n-1}/\sqrt{n}} < c\}$ con c verificando que: $P(t_{n-1} < c) = \alpha$.

3.
$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 (\mu_1 \neq \mu_0) \end{cases} \tag{55}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid\mid \frac{\overline{X} - \mu_0}{S_{n-1}/\sqrt{n}} \mid\mid > c\}$ con c verificando que: $P(\mid t_{n-1} \mid > c) = \alpha$.

Ejemplo 3.9
Supongamos que nos dicen que la ley de probabilidad de la variable que mide el tiempo de convergencia de un algoritmo estocástico de optimización sigue una ley normal. Extraemos una muestra de tamaño 20, y obtenemos $\overline{X} = 58$ sg. con $S_{n-1} = 10$ sg. Nos preguntamos si estos datos son evidencia suficiente para refutar a nivel $\alpha = 0.05$ la hipótesis de que el tiempo de ejecución esperado es de 60 sg. frente a la alternativa de que dicho tiempo de ejecución esperado es de 55 sg.

El test planteado es:

$$\begin{cases} H_0 : \mu = 60 \\ H_1 : \mu = 55 \end{cases} \tag{56}$$

Según el teorema 3.4, la región de rechazo es

$$R.R. = \{(x_1, \ldots, x_n) \mid \frac{\overline{X} - \mu_0}{S_{n-1}/\sqrt{n}} < c\} \tag{57}$$

con c verificando que: $P(t_{19} < c) = 0.05$. De ahí que $c = -1.73$. Al verificarse que $\frac{58-60}{10/\sqrt{20}} > -1.73$, la muestra no pertenece a la región de rechazo, de ahí que la decisión que se deriva del test sea la de aceptar H_0.

Al igual que en el caso en el que la varianza era conocida, podemos tambien en este caso, planteranos por tests de hipótesis con hipótesis alternativas no reducidas a un punto.

6 Test de hipótesis para la igualdad de esperanzas matemáticas de dos distribuciones normales

En este apartado estudiaremos los tests de hipótesis que surgen al plantearse por la igualdad de esperanzas matemáticas de dos distribuciones normales. Plateamos dos casos dependiendo de si las muestras a partir de las cuales se ha de tomar la decisión son *muestras independientes* o *muestras apareadas*.

6.1 Muestras independientes

Vamos a introducir, apoyándonos en un ejemplo, el concepto de *muestras independientes*. Conviene aclarar que el adjetivo independiente se usa en este caso con un significado distinto al que se le ha venido dando cuando habláblamos de variables aleatorias independientes. Tal y como se verá en el ejemplo, la idea que subyace al concepto de muestras independientes se relaciona con que dichas muestras no están relacionadas.

Ejemplo 3.10
Supongamos dos heurísticos de búsqueda que denotaremos respectivamente por A y B. Igualmente las variables aleatorias X_A y X_B van a denotar el tiempo expresado en segundos hasta que respectivamente los heurísticos A y B convergen. Extraemos una muestra aleatoria de X_A de tamaño n_A, en la cual hemos obtenido como media aritmética $\overline{X}_A$ y varianza muestral $S^2_{n_A}$. Análogamente para el heurístico B se han obtenido $\overline{X}_B$ y $S^2_{n_B}$. A partir de dicha información queremos testar la hipótesis de que la esperanza matemática de la variable tiempo de respuesta para el heurístico A es igual a la del heurístico B.

Las dos muestras anteriores se dice que son muestras independientes ya que el i–ésimo valor obtenido con el heurístico A no está relacionado con el i–ésimo valor obtenido con el heurístico B. Es decir que el orden en el que se colocan –véase la tabla 3.3– los valores de ambas muestras es irrelevante.

Tabla 3.3: Ejemplo de muestras independientes, siendo $n_A < n_B$.

	heurístico A	heurístico B
1	x_1^A	x_1^B
2	x_2^A	x_2^B
...	...	...
n_A	$x_{n_A}^A$	$x_{n_A}^B$
n_A+1		$x_{n_A+1}^B$
...		...
n_B		$x_{n_B}^B$

Teorema 3.5
Sean X_A y X_B dos variables aleatorias siguiendo distribuciones normales con leyes $\mathcal{N}(\mu_A, \sigma_A)$ y $\mathcal{N}(\mu_A, \sigma_A)$ respectivamente, siendo σ_A y σ_B dos parámetros desconocidos pero iguales.[1] De las variables anteriores se extraen dos muestras independientes de tamaños respectivos n_A y n_B, obteniéndose para la primera variable el valor $\overline{X}_A$ como media aritmética y $S^2_{n_A-1}$ como cuasivarianza muestral. Análogamente para la segunda variables se han obtenido $\overline{X}_B$ y $S^2_{n_B-1}$.

La región de rechazo $R.R.$ para el test de hipótesis de igualdad de esperanzas matemáticas siguiente:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases} \tag{58}$$

viene dado por

$$R.R. = \{(x_1^A, \ldots, x_{n_A}^A, x_1^B, \ldots, x_{n_B}^B) \mid \frac{\overline{X}_A - \overline{X}_B}{S_{n_A+n_B-2}\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} \mid > c\} \tag{59}$$

con c verificando que: $P(\mid t_{n_A+n_B-2} \mid > c) = \alpha$ y $S^2_{n_A+n_B-2} = \frac{(n_A-1)S^2_{n_A-1}+(n_B-1)S^2_{n_B-1}}{n_A+n_B-1}$.

Ejemplo 3.11
Supongamos que dos muestras independientes extraidas de dos variables normales X_A y X_B de las que no se conocen las varianzas respectivas, pero se asume que son iguales, han proporcionado los siguientes datos:

$$\begin{array}{lll} n_A = 13, & \overline{X}_A = 45.6 & S^2_{n_A-1} = 1.8 \\ n_B = 10, & \overline{X}_B = 36.3 & S^2_{n_B-1} = 3.6 \end{array}$$

Se trata de testar la hipótesis:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases} \tag{60}$$

a un nivel de significación de $\alpha = 0.05$.

Para ver si los datos de las dos muestras independientes nos llevan a aceptar o rechazar la hipótesis nula, debemos de determinar el valor de c, así como el valor de la expresión:

$$\mid \frac{\overline{X}_A - \overline{X}_B}{S_{n_A+n_B-2}\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} \mid \tag{61}$$

El valor de c que verifica que: $P(\mid t_{13+10-2} \mid > c) = 0.05$ es $c = 0.26$.

La anterior expresión vale:

[1] Esto quiere decir que se ha testado la hipótesis nula de igualdad de varianzas y no se ha encontrado evidencia suficiente como para rechazarla.

$$| \frac{45.6 - 36.3}{\sqrt{\frac{(13-1)1.8+(10-1)3.6}{13+10-2}}\sqrt{\frac{1}{13}+\frac{1}{10}}} |> 0.26$$

De ahí que la decisión adoptada es la de rechazar la hipótesis nula.

6.2 Muestras apareadas

Veamos con un ejemplo el concepto de muestras apareadas.

Ejemplo 3.12
Supongamos que tenemos dos algoritmos de ordenación distintos que denotamos respectivamente por A y B. Supongamos asimismo que la rapidez de dichos algoritmos depende de la ristra de valores a ordenar. Denotamos por X_A la variable aleatoria midiendo el tiempo que el algoritmo A necesita para llevar a cabo dicha ordenación. Dicho tiempo es aleatorio, aunque el algoritmo A sea determinista, ya que depende de los valores de la ristra que se trata de ordenar. Análogamente X_B denotará la variable aleatoria midiendo el tiempo relativo al algoritmo B. Con objeto de comparar ambos algoritmos, los ejecutamos con n ristras de valores distintos, reflejando los valores obtenidos en la tabla 3.4. El valor muestral x_i^A denota para el algoritmo A el tiempo necesario para ordenar la ristra i. De manera similar interpretamos x_i^B en relación con el algoritmo B.

Decimos que las muestras son apareadas, ya que los valores que están en la i–ésima fila de la tabla 3.4 se refieren a la misma ristra –la i–ésima– a ordenar. Por tanto el orden de ambas columnas debe mantenerse en cuanto que los valores de cada fila se refieren al mismo conjunto de datos.

Tabla 3.4: Ejemplo de muestras apareadas.

	heurístico A	heurístico B
1	x_1^A	x_1^B
2	x_2^A	x_2^B
...	...	...
...	...	...
...	...	...
i	x_i^A	x_i^B
...	...	...
...	...	...
...	...	...
n	x_n^A	x_n^B

Teorema 3.6
Sean X_A y X_B dos variables aleatorias siguiendo distribuciones normales con leyes $\mathcal{N}(\mu_A, \sigma_A)$ y $\mathcal{N}(\mu_B, \sigma_B)$ respectivamente. De las variables anteriores se extraen dos muestras apareadas de tamaño n, obteniéndose para la primera variable $\overline{X}_A$ como media aritmética y $S^2_{n_A-1}$ como cuasi-varianza muestral. Análogamente para la segunda variable se han obtenido $\overline{X}_B$ y $S^2_{n_B-1}$.

La región de rechazo $R.R.$ para el test de hipótesis de igualdad de esperanzas matemáticas siguiente:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases} \tag{62}$$

viene dado por

$$R.R. = \{(x_1^A, \ldots, x_n^A, x_1^B, \ldots, x_n^B) \mid \frac{\overline{X}_A - \overline{X}_B}{S_{AB,n-1}/\sqrt{n}} |> c\} \tag{63}$$

con c verificando que: $P(|\ t_{n-1}\ | > c) = \alpha$ y $S^2_{AB,n-1} = \frac{1}{n-1}\sum_{i=1}^{n}((x_i^A - x_i^B) - (\overline{X}^A - \overline{X}^B))^2$.

Ejemplo 3.13
A partir de los valores registrados en la tabla 3.5, los cuales corresponden a dos muestras relacionadas de tamaño 10 extraidas de dos variables normales X_A y X_B con esperanzas matemáticas desconocidas denotadas respectivamente por μ_A y μ_B, se trata de efectuar el siguiente contraste:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases} \tag{64}$$

con un nivel de significación de $\alpha = 0.05$.

Tabla 3.5: Ejemplo de muestras apareadas.

	X_A	X_B
1	140	145
2	165	150
3	160	150
4	160	160
5	175	170
6	190	175
7	170	160
8	175	165
9	155	145
10	160	170

A partir de las tablas de la t_9, calcularemos c, como el punto verificando: $P(|\ t_9\ | > c) = 0.05$. Obtenemos que $c = 2.262$.

Por otra parte $\overline{X}_A - \overline{X}_B = 6$. Además: $S^2_{AB,n-1} = \frac{1}{9}[(-5-6)^2 + (15-6)^2 + \ldots + (-10-6)^2] = 71.11$.

Como $|\ \frac{6}{\sqrt{71.11}/\sqrt{10}}\ | = 2.25 < 2.262$ concluimos que las dos muestras relacionadas pertenecen a la región de aceptación. Dicho de otra manera, las diferencias encontradas en las muestras extraidas de ambas distribuciones, no han resultado ser lo suficientemente grandes como para que reciban el calificativo de diferencias estadísticamente significativas.

7 Test de hipótesis para el parámetro p de una distribución de Bernouilli

Los tests de hipótesis que se presentarán en este apartado son tambien conocidos como tests de hipótesis para una proporción. Uno de ellos ha sido ya obtenido como aplicación del lema de Neyman–Pearson.

Teorema 3.7
Sea X una variable aleatoria de Bernouilli con parámetro p desconocido. A partir de una muestra de tamaño n extraida de X, la región de rechazo de nivel α, obtenida al aplicar el test de la razón de verosimilitud para:

1.
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 (p_1 > p_0) \end{cases} \tag{65}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \sum_{i=1}^{n} X_i > c\}$ con c verificando $P(B(n, p_0) \geq \lceil c \rceil + 1) = \alpha$.

2.
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 (p_1 < p_0) \end{cases} \tag{66}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \sum_{i=1}^{n} X_i < c\}$ con c verificando $P(B(n, p_0) \leq \lceil c \rceil + 1) = \alpha$.

3.
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \tag{67}$$

es: $R.R. = \{(x_1, \ldots, x_n) \mid \sum_{i=1}^{n} X_i < c_1 o \sum_{i=1}^{n} X_i > c_2\}$ con c_1 y c_2 verificando que $c_1 < c_2$ y $P(B(n, p_0) \leq \lceil c_1 \rceil - 1) + P(B(n, p_0) \geq \lceil c_2 \rceil + 1) = \alpha$.

Al igual que en el apartado 3.5, tambien en este caso nos podemos plantear por tests de hipótesis en los que la alternativa sea más genérica que la expresada en los dos primeros casos anteriores. Por ejemplo:

1.
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases} \tag{68}$$

2.
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases} \tag{69}$$

8 Test de hipótesis para la igualdad de esperanzas matemáticas de dos distribuciones de Bernouilli

El nombre más habitual para este tipo de tests es el de tests de comparación de dos proporciones. Al igual que en el apartado 3.6 contemplaremos dos casos en función de que las muestras sean independientes o de que las muestras sean apareadas.

8.1 Muestras independientes

En este caso vamos a presentar un test aproximado ya que en su construcción se utiliza el teorema central del límite.

Teorema 3.8
Sean X_A y X_B dos variables aleatorias siguiendo distribuciones de Bernouilli de parámetros p_A y p_B respectivamente, siendo ambos desconocidos. De las variables anteriores se extraen dos muestras independientes de tamaños respectivos n_A y n_B. Se denotan por $\hat{p}_A$ y $\hat{p}_B$ a las frecuencias relativas de éxito obtenidas para las muestras de tamaños respectivos n_A y n_B. Es decir $\hat{p}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} X_j^A$ y $\hat{p}_B = \frac{1}{n_B} \sum_{i=1}^{n_B} X_j^B$. Denotemos por $\hat{p}$ a la frecuencia relativa de éxito obtendia en la muestra global de tamaño $n_A + n_B$. Es decir $\hat{p} = \frac{1}{n_A + n_B} (\sum_{i=1}^{n_A} X_i^A + \sum_{i=1}^{n_B} X_i^B) = \frac{n_A \hat{p}_A + n_B \hat{p}_B}{n_A + n_B}$.

La región de rechazo $R.R.$ para el test de hipótesis de igualdad de esperanzas matemáticas de dos poblaciones de Bernouilli siguiente:

$$\begin{cases} H_0 : p_A = p_B \\ H_1 : p_A \neq p_B \end{cases} \tag{70}$$

viene dado por:

$$R.R. = \{(x_1^A, \ldots, x_{n_A}^A, x_1^B, \ldots, x_{n_B}^B) \mid \frac{\mid \hat{p_A} - \hat{p_B} \mid - \frac{1}{4}(\frac{1}{n_A} + \frac{1}{n_B})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A} + \frac{1}{n_B})}} > c\} \tag{71}$$

con c verificando que: $P(\mid \mathcal{N}(0,1) \mid > c) = \alpha$.

El término $\frac{1}{4}(\frac{1}{n_A} + \frac{1}{n_B})$ actúa como un factor corrector de continuidad y viene justificado por la utilización de la distribución normal como distribución que aproxima a la verdadera distribución del estadístico a partir del cual se construye el test.

Ejemplo 3.14
Supongamos que queramos testar la hipótesis nula de igualdad en las probabilidades de alcanzar el óptimo con dos algoritmos estocásticos de optimización que denotamos por A y B respectivamente. Con el algoritmo A obtenemos que en 90 ejecuciones del mismo, hemos alcanzado el óptimo 21 veces, mientras que con el algoritmo B de 104 ejecuciones se ha alcanzado el óptimo en 32 de ellas. Nos preguntamos si las diferencias entre las proporciones de éxito encontradas en las muestras son lo suficientemente grandes como para rechazar la hipótesis nula.

Por los datos anteriores tenemos que:

$$X_A \to B(1, p_A); X_B \to B(1, p_B) \tag{72}$$

A partir de dos muestras independientes de tamaños respectivos $n_A = 90$ y $n_B = 104$, hemos obtenido $\hat{p}_A = \frac{21}{90}$ y $\hat{p}_B = \frac{32}{104}$. Además $\hat{p} = \frac{21+32}{90+104} = \frac{53}{194}$.

Para ver si la muestra global de tamaño 194 pertenece a la región de rechazo o a la región de aceptación debemos de comparar el valor de $\frac{|\hat{p}_A - \hat{p}_B| - \frac{1}{4}(\frac{1}{n_A}+\frac{1}{n_B})}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_A}+\frac{1}{n_B})}}$ con $c = 1.96$, valor este último obtenido en las tablas de una distribución $\mathcal{N}(0,1)$.

Al verificarse que:

$$\frac{\mid \frac{21}{90} - \frac{32}{104} \mid -\frac{1}{4}(\frac{1}{90} + \frac{1}{104})}{\sqrt{\frac{53}{194}\frac{141}{190}(\frac{1}{90} + \frac{1}{104})}} = 1.07 < 1.96 \tag{73}$$

concluimos que la muestra global pertenece a la región de aceptación y se decide aceptar la hipótesis nula, ya que las diferencias encontradas entre las dos proporciones no llegan a ser lo suficientemente grandes como para ser consideradas estadísticamente significativas.

8.2 Muestras apareadas

En este punto se presentará el denominado test de McNemar.

Teorema 3.9
Sean X_A y X_B dos variables aleatorias siguiendo distribuciones de Bernouilli de parámetros p_A y p_B respectivamente, siendo ambos desconocidos. De las variables anteriores se extraen dos muestras apareadas de tamaño n. La región de rechazo $R.R.$ del test de McNemar para el test de hipótesis siguiente:

$$\begin{cases} H_0 : p_A = p_B \\ H_1 : p_A \neq p_B \end{cases} \tag{74}$$

viene dado por:

$$R.R. = \{(x_1^A, \ldots, x_{n_A}^A, x_1^B, \ldots, x_{n_B}^B) \mid \frac{\mid n_{12} - n_{21} \mid -0.5}{\sqrt{n_{12} + n_{21}}} > c\} \tag{75}$$

con c verificando que: $P(\mid \mathcal{N}(0,1) \mid > c) = \alpha$, siendo:
n_{12} el número de casos, de los n, en los que la variable X_A toma el valor éxito, y a la vez, la variable X_B toma el valor fracaso, y n_{21} el número de casos, de los n, en los que la variable X_A toma el valor fracaso, y a la vez, la variable X_B toma el valor éxito.

Ejemplo 3.15
Supongamos que queramos comparar las probabilidades de que dos páginas web A y B de un mismo portal sean visitadas. Para ello recogemos la información de 125 casos. Dicha información se muestra en la tabla 3.5. La variable X_A denota el éxito (página A visitada) o el fracaso en relación con la primera de las páginas. De mánera análoga la variable X_B recoge los valores en relación con la página B.

Tabla 3.5: Datos para la comparación de dos proporciones en muestras apareadas.

		X_A		
		éxito	fracaso	
X_B	éxito	$n_{11} = 27$	$n_{21} = 35$	62
	fracaso	$n_{12} = 43$	$n_{22} = 20$	63
		70	55	125

Se trata de comparar $\frac{|n_{12}-n_{21}|-0.5}{\sqrt{n_{12}+n_{21}}}$ con $c = 1.96$. Al verificarse que: $\frac{|43-35|-0.5}{\sqrt{43+35}} < 1.96$, concluimos que las diferencias entre las proporciones, $\hat{p}_A = \frac{70}{125}$ y $\hat{p}_B = \frac{62}{125}$ no son lo suficientemente grandes como para rechazar la hipótesis nula.

9 Test de Kolmogoroff–Smirnov

Tanto en el tema 1 como en el tema 2 e incluso en este tema hemos venido haciendo suposiciones acerca de la ley de probabilidad que seguia una determinada variable. Mientras que en algunos casos de distribuciones discretas (por ejemplo en distribuciones de Bernouilli, distribuciones binomiales, distribuciones binomial negativas, hipergeométricas, polinomiales, ...) el experimento aleatorio realizado determinaba la distribución de probabilidad a utilizar, en los casos de variables aleatorias continuas necesitamos contrastar si los datos de la muestra siguen un determinado modelo probabilístico.

Una forma intuitiva de hacer esto consiste en construir el histograma de frecuencias relativas a partir de los datos muestrales y posteriormente tratar de encontrar entre las curvas de función de densidad correspondientes a los distintos modelos de variables aleatorias continuas, aquel que proporcione un mejor ajuste. Si bien este procedimiento tiene como ventaja fundamental el que se trata de una comparación basada en la intuición, su mayor desventaja consiste en que el proceso de discretizar la variable continua para así obtener el histograma es totalmente arbitrario. Además se necesita un criterio objetivo –por ejemplo un estadístico del cual conozcamos su ley de probabilidad– con el que medir la distancia entre el histograma de frecuencias relativas y la función de densidad a seleccionar.

En este apartado vamos a presentar un resultado conocido como test de Kolmogoroff-Smirnov que va a permitir llevar a cabo dicha comparación entre los datos muestrales y el modelo que se propone.

Un concepto en el que se basa el test de Kolmogoroff-Smirnov es en el de función de distribución muestral que definimos a continuación.

Definición 3.6
Dada una muestra aleatoria simple de tamaño n, $x_1, \ldots, x_n$ extraida de la variable aleatoria X, denotamos por $x_{1:n}, \ldots, x_{n:n}$ a la muestra anterior ordenada[2]. La *función de distribución muestral* asociada a la muestra anterior se denota por $F_n(x)$ y se define como:

[2]Es decir se verifica: $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$

$$F_n(x) = \begin{cases} 0 & \text{si } x < x_{1:n} \\ \dots & \dots \\ \frac{i}{n} & \text{si } x_{i:n} \leq x \leq x_{i+1:n} (i = 1, ..., n) \\ \dots & \dots \\ 1 & \text{si } x \geq x_{n:n} \end{cases} \tag{76}$$

Por tanto la función de distribución muestral mide en cada punto el porcentaje de casos que en la muestra anterior toman valores menores o iguales que dicho punto.

Veamos, en el resultado siguiente, como el test de Kolmogoroff-Smirnov contrasta el ajuste de una ley de probabilidad a unos datos muestrales, calculando la máxima diferencia entre la función de distribución de la ley de distribución teórica propuesta y la función de distribución muestral. Dicha máxima diferencia deberá ser menor que un determinado umbral para poder aceptar que la ley de probabilidad teórica propuesta constituye un buen modelo para los datos. Un aspecto importante a resaltar es que el umbral anterior tan sólo depende del tamaño de la muestra y del nivel de significación al que se quiere construir el test, siendo independiente de la ley de probabilidad teórica propuesta.

Teorema 3.10
Sea X una variable aleatoria con ley de probabilidad desconocida. Extraemos una muestra aleatoria de X de tamaño n. La región de rechazo a nivel α para el siguiente test de hipótesis:

$$\begin{cases} H_0: & X \text{ tiene como función de distribución } F_X^0(x) \\ H_1: & X \text{ no tiene como función de distribución } F_X^0(x) \end{cases} \tag{77}$$

es de la forma siguiente:

$$R.R. = \{(x_1, \dots, x_n) \mid \max_{x_i} \mid F_n(x_i) - F_X^0(x) \mid > c\} \tag{78}$$

donde $F_n(x)$ representa la función de distribución muestral y el valor de c se calcula como el valor que en las tablas del estadístico de Kolmogoroff-Smirnov deja a su derecha un área de α.

Tal y como se ve en el resultado anterior el estadístico de Kolmogoroff-Smirnov no depende de la ley de probabilidad teórica.

En el caso de que bajo la hipótesis nula no se especifiquen los parámetros de la ley de probabilidad teórica, estos deberán de estimarse por sus estimadores máximo verosímiles, y deberíamos de aplicar una variante del test de Kolmogoroff-Smirnov conocida como Massey-Lillifors.

La figura 3.2 muestra de forma gráfica como calcular el estadístico de Kolmogoroff-Smirnov.

Figura 3.2: Diferencia entre la función de distribución teórica bajo H_0, $F_X^0(x)$, y la función de distribución muestral, $F_n(x)$.

Ejemplo 3.16
Supongamos que nos planteamos el testar la hipótesis de que los datos recogidos en la tabla 3.6, que recogen el tiempo de 20 ejecuciones de un algoritmo estocástico, siguen una distribución normal $\mathcal{N}(60, 10)$.

Tabla 3.6: Tiempo de 20 ejecuciones de un algoritmo estocástico de optimización.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
68	64	81	58	70	65	62	71	59	60	61	73	78	46	49	75	67	77	50	69

Según el teorema anterior debemos de:

(i) ordenar en orden creciente los valores de la muestra, obteniendo: $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$
(ii) calcular $F_n(x_{i:n})$
(iii) calcular $F_X^0(x)$
(iv) encontrar la mayor diferencia entre $F_n(x_{i:n})$ y $F_X^0(x)$
(v) comparar dicha máxima diferencia con el valor que en las tablas del estadístico de Kolmogoroff-Smirnov deja a su derecha un área de α.

La tabla 3.7 muestra los valores que la función de distribución muestral y la función de distribución teórica toman en los datos muestrales.

Tabla 3.7: Función de distribución muestral y teórica bajo H_0.

	$x_{1:20}$	$x_{2:20}$	$x_{3:20}$	$x_{4:20}$	$x_{5:20}$	$x_{6:20}$	$x_{7:20}$	$x_{8:20}$	$x_{9:20}$	$x_{10:20}$
$x_{i:n}$	46	49	50	58	59	60	61	62	64	65
$F_n(x_{i:n})$	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{5}{20}$	$\frac{6}{20}$	$\frac{7}{20}$	$\frac{8}{20}$	$\frac{9}{20}$	$\frac{10}{20}$
$F_X^0(x)$	0.09	0.14	0.16	0.43	0.47	0.50	0.53	0.57	0.65	0.69
	$x_{11:20}$	$x_{12:20}$	$x_{13:20}$	$x_{14:20}$	$x_{15:20}$	$x_{16:20}$	$x_{17:20}$	$x_{18:20}$	$x_{19:20}$	$x_{20:20}$
$x_{i:n}$	67	68	69	70	71	73	75	77	78	81
$F_n(x_{i:n})$	$\frac{11}{20}$	$\frac{12}{20}$	$\frac{13}{20}$	$\frac{14}{20}$	$\frac{15}{20}$	$\frac{16}{20}$	$\frac{17}{20}$	$\frac{18}{20}$	$\frac{19}{20}$	$\frac{20}{20}$
$F_X^0(x)$	0.75	0.78	0.81	0.84	0.86	0.90	0.93	0.95	0.96	0.98

Al comparar $\max_{x_{i:n}} \mid F_n(x_{i:n}) - F_X^0(x_{i:n}) \mid = \mid F_{20}(x_{4:20}) - F_X^0(x_{4:20}) \mid = 0.23$ con $c = 0.26$ obtenemos que la muestra pertenece a la región de aceptación por lo que se decide no rechazar la hipótesis nula.

10 Test de hipótesis basados en la distribución chi–cuadrado

En este apartado vamos a presentar tres test de hipótesis basados en la distribución chi–cuadrado. El primero de los tests se utilizará para testar la independencia entre dos variables nominales u ordinales. El segundo servirá para plantearse por la homogeneidad de dos o más distribuciones polinomiales. Este segundo test, aunque utiliza el mismo estadístico que el que sirve para testar la independencia, presenta diferencias conceptuales en cuanto a las hipótesis nula y alternativa testadas. El tercer test presentado sirve para contrastar el ajuste de una distribución teórica discreta con respecto de unos datos muestrales, y puede verse como una alternativa al test de Kolmogoroff–Smirnov para variables discretas.

10.1 Test de independencia entre dos variables aleatorias discretas

Vamos a ilustrar la utilización del test de independencia por medio de un ejemplo que hace referencia a los resultados obtenidos con tres algoritmos heurísticos estocásticos de optimización.

Ejemplo 3.17
Las tablas 3.8 y 3.9 recogen respectivamente los resultados obtenidos al comparar los heurísticos estocásticos de optimización A y B (tabla 3.8) y A y C (tabla 3.9). Las variables aleatorias X, Y, Z se relacionan respectivamente con los resultados obtenidos por los algoritmos A, B y C. Para cualesquiera de las tres variables aleatorias anteriores un valor de 0 denota que el algoritmo alcanza el óptimo global del problema o bien no alcanza el óptimo global, pero el valor alcanzado queda a menos de un 5 por ciento de distancia con respecto a dicho óptimo global. El valor 1 significa que el heurístico queda a una distancia superior al 5 por ciento del valor óptimo.

Tabla 3.8: Resultados obtenidos con los heurísticos estocásticos A y B.

		X		
		0	1	
Y	0	95	5	100
	1	105	795	900
		200	800	1000

Tabla 3.9: Resultados obtenidos con los heurísticos estocásticos A y C.

		X		
		0	1	
Z	0	81	319	400
	1	119	481	600
		200	800	1000

El objetivo de los tests de hipótesis de independencia será el decidir acerca de la independencia entre las dos variables aleatorias consideradas. Mientras que en el ejemplo recogido en la tabla 3.9, independientemente de que el valor de la variable X sea 0 o 1, la proporción de los valores de Z que son 0 o 1 permanece cercana a $\frac{4}{6}$ ($\frac{81}{119}$ cuando X vale 0, y $\frac{319}{481}$ cuando X vale 1)[3] la situación que se manifiesta en la tabla 3.8 es totalmente distinta. En esta última tabla, cuando el valor de X vale 0, la proporción entre los valores de Y es de $\frac{95}{105}$, mientras que cuando X vale 1, dicha proporción se reduce a $\frac{5}{795}$. Análogamente, cuando Y vale 0, la proporción entre los valores de X es de $\frac{95}{5}$, mientras que cuando Y toma el valor 1, dicha proporción es de $\frac{105}{795}$.

De manera intuitiva podemos decir que los resultados de la tabla 3.9 apoyan la independencia entre X y Z, mientras que los datos de la tabla 3.8, están reflejando una situación de dependencia

[3]Nótese que tambien ocurre que independientemente de que el valor de la variable Z sea 0 o 1, la proporción de los valores de X que son 0 o 1, permanece cercana a $\frac{2}{8}$ ($\frac{81}{319}$ cuando Z vale 0, y $\frac{119}{481}$ cuando Z vale 1)

entre las variables X e Y.

A continuación introduciremos un test que sirve para llevar a cabo el contraste de independencia, aunque previamente necesitamos definir el concepto de tabla de contingencia.

DEFINICIÓN 3.7
Sean X e Y dos variables aleatorias discretas. Supongamos que X puede tomar f posibles valores que denotamos por $x^1, \dots x^i, \dots, x^f$ mientras que los c posibles valores que puede llegar a tomar la variable Y los denotamos por $y^1, \dots y^j, \dots, y^c$. Extraemos una muestra aleatoria simple de tamaño n de la variable bidimensional (X, Y) y denotamos por O_{ij} el número de casos, de entre los n, en los que la variable X toma su i-ésimo valor, x^i, y a la vez la variable Y toma su j-ésimo valor y^j. El símbolo $n_{i.}$ denota el número de casos en los que la variable X toma su i-ésimo valor, mientras que $n_{.j}$ denota el número de casos en los que la variable Y toma su j-ésimo valor. Se tiene por tanto que: $n_{i.} = \sum_{j=1}^{c} O_{ij}$ y $n_{.j} = \sum_{i=1}^{f} O_{ij}$. Además se verifica que:

$$n = \sum_{i=1}^{f} n_{i.} = \sum_{j=1}^{c} n_{.j} = \sum_{i=1}^{f} \sum_{j=1}^{c} O_{ij}. \tag{79}$$

La tabla 3.10 representa la situación descrita.

Tabla 3.10: Tabla de contingencia en un test de independencia.

		Y						
		y^1	y^2	...	y^j	...	y^c	
	x^1	O_{11}	O_{12}	...	O_{ij}	...	O_{1c}	$n_{1.}$
	x^2	O_{21}	O_{22}	...	O_{2j}	...	O_{2c}	$n_{2.}$
	...	...	...	...	...	...	...	...
X	x^i	O_{i1}	O_{i2}	...	O_{ij}	...	O_{ic}	$n_{i.}$
	...	...	...	...	...	...	...	...
	x^f	O_{f1}	O_{f2}	...	O_{fj}	...	O_{fc}	$n_{f.}$
		$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	n

TEOREMA 3.11
Sean dos variables aleatorias discretas X e Y con posibles valores: $x^1, \dots x^i, \dots, x^f$ e $y^1, \dots y^j, \dots, y^c$ respectivamente. A partir de una muestra aleatoria simple de tamaño n: $\{(x_1, y_1), \dots, (x_n, y_n)\}$ extraida de la variable aleatoria bidimensional (X, Y), el siguiente test de hipótesis:

$$\begin{cases} H_0: & X \text{ e } Y \text{ son variables aleatorias independientes} \\ H_1: & X \text{ e } Y \text{ no son variables aleatorias independientes} \end{cases} \tag{80}$$

tiene como región de rechazo:

$$R.R. = \{(x_1, y_1), \dots, (x_n, y_n) | \sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \geq c\} \tag{81}$$

donde:

- O_{ij} denota el número de casos observados en la muestra de tamaño n en los que X vale x^i y a la vez Y vale y^j
- E_{ij} denota el número de casos esperados bajo la hipótesis nula de independencia entre X e Y en los que X debe valer x^i y a la vez Y debe tomar el valor y^j. La fórmula para calcular E_{ij} como veremos más adelante es la siguiente: $E_{ij} = \frac{n_{i.} n_{.j}}{n}$
- $n_{i.} = \sum_{j=1}^{c} O_{ij}$
- $n_{.j} = \sum_{i=1}^{f} O_{ij}$
- $n = \sum_{i=1}^{f} \sum_{j=1}^{c} O_{ij}$

- c es una constante tal que: $P(\chi^2_{(f-1)(c-1)} \geq c) = \alpha$.

La idea subyacente al test consiste en comparar, para cada celdilla, el número de casos observados, O_{ij}, con el número de casos esperados, E_{ij}, según el modelo de independencia. En realidad dicho modelo de independencia establecido en la hipótesis nula puede expresarse para $i = 1, \ldots, f; j = 1, \ldots, c$ como:

$$H_0 : p_{ij} = p_{i.}p_{.j} \tag{82}$$

denotando $p_{ij} = P(X = x^i, Y = y^j)$, $p_{i.} = P(X = x^i)$, $p_{.j} = P(Y = y^j)$.

Bajo dicha hipótesis de independencia, E_{ij}, puede expresarse como:

$$E_{ij} = np_{i.}p_{.j} = n\frac{n_{i.}}{n}\frac{n_{.j}}{n} = \frac{n_{i.}n_{.j}}{n} \tag{83}$$

Ejemplo 3.18
Si queremos testar la hipótesis nula de independencia entre las variables X e Y del ejemplo 3.17, tenemos que:

$$\begin{cases} H_0 : & X \text{ e } Z \text{ son variables aleatorias independientes} \\ H_1 : & X \text{ e } Z \text{ no son variables aleatorias independientes} \end{cases} \tag{84}$$

y debemos de comparar el valor de $c = 3.84$ (ya que $P(\chi^2_{(2-1)(2-1)} \geq 3.84) = 0.05$) con:

$$\sum_{i=1}^{f}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} =$$

$$\frac{(81-80)^2}{80} + \frac{(319-320)^2}{320} + \frac{(119-120)^2}{120} + \frac{(481-480)^2}{480} =$$
$$0.0125 + 0.003125 + 0.0083 + 0.00208 = 0.026005 < 3.841 \tag{85}$$

de ahí que concluyamos que no podemos rechazar la hipótesis nula que dice que las variables X y Z son independientes.

10.2 Test de homogeneidad

Este test de homogeneidad presenta unas características análogas al expuesto en el apartado anterior, el cual tenía como objetivo el testar la independencia entre dos variables aleatorias discretas. La diferencia básica radica en que en este caso se plantea el estudiar si el comportamiento de una variable puede, o no, considerarse homogéneo en distintos grupos.

Ejemplo 3.19
Supongamos que estemos interesados en estudiar si el comportamiento del algoritmo heurístico estocástico que denotamos por A es homogéneo en distintas instancias del problema del agente viajero. Para ello consideramos 4 instancias de dicho problema de optimización combinatorial, las cuales varían en el número de ciudades a visitar, siendo sus valores 20, 50, 100 y 500. Denotamos respectivamente por TSP20, TSP50, TSP100 y TSP500 a los 4 ejemplos del problema del agente viajero considerados. Por otra parte la variable aleatoria X recoge el resultado de aplicar el heurístico A a los problemas anteriores. Al igual que en los dos ejemplos anteriores el valor 0 indica que el algoritmo A alcanza una solución que dista del óptimo global (conocido) menos que el 5 por ciento, mientras que el valor 1 en la variable X indica que la solución alcanzada por el heurístico A supera en un 5 por ciento el valor del óptimo global.

La tabla 3.11 recoge los resultados alcanzados con 100 ejecuciones del algoritmo A en cada uno de los 4 ejemplos del problema del agente viajero considerados.

Tabla 3.11: Resultados obtenidos con el heurístico estocástico A en cuatro instancias del problema del agente viajero.

	X		
	0	1	
TSP20	94	6	100
TSP50	81	19	100
TSP100	73	27	100
TSP500	52	48	100
	300	100	400

La pregunta que nos hacemos es si podemos considerar que el comportamiento del heurístico estocástico A ha sido homogéneo o no en las 4 instancias del problema del agente viajero. Al observar los datos de la tabla 3.11 podemos ver que a medida que aumenta la complejidad del problema la proporción de veces que el algoritmo A obtiene soluciones que no se alejan más del 5 por ciento del óptimo global, va reduciéndose. Veremos, en el resultado que presentamos a continuación, como decidir si las diferencias en dichas proporciones pueden o no ser consideradas estadísticamente significativas.

Teorema 3.12
Sea una variable aleatoria discreta, X, con f posibles valores: $x^1, \dots x^i, \dots, x^f$. Sean c grupos o poblaciones, que denotamos por: $GR^1, \dots GR^j, \dots, GR^c$, en los cuales se experimenta con la variable aleatoria X. A partir de una muestra global de tamaño n, constituida por c muestras de tamaños respectivos $n_{.1}, \dots, n_{.j}, \dots, n_{.c}$ (cada una de las cuales está asociada a un grupo) –veáse tabla 3.12– , nos planteamos por el test de hipótesis relativo a la homogeneidad en el comportamiento de la variable aleatoria X en cada uno de los c grupos. Es decir:

$$\begin{cases} H_0: & X \text{ es homogénea en } GR^1, \dots GR^j, \dots, GR^c \\ H_1: & X \text{ no es homogénea en } GR^1, \dots GR^j, \dots, GR^c \end{cases} \tag{86}$$

La región de rechazo para dicho test está constituida por aquellas muestras en las que se verifica:

$$\sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \geq c \tag{87}$$

donde O_{ij}, E_{ij} y c se definen de la misma manera que el teorema 3.11.

Tabla 3.12: Tabla de contingencia en un test de homogeneidad.

		GR^1	GR^2	...	GR^j	...	GR^c	
	x^1	O_{11}	O_{12}	...	O_{ij}	...	O_{1c}	$n_{1.}$
	x^2	O_{21}	O_{22}	...	O_{2j}	...	O_{2c}	$n_{2.}$
	...	...	...	...	...	...	...	...
X	x^i	O_{i1}	O_{i2}	...	O_{ij}	...	O_{ic}	$n_{i.}$
	...	...	...	...	...	...	...	...
	x^f	O_{f1}	O_{f2}	...	O_{fj}	...	O_{fc}	$n_{f.}$
		$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.c}$	n

Ejemplo 3.20
Veamos la aplicación del teorema anterior a los datos recogidos en la tabla 3.11. Es decir nos planteamos si el comportamiento del heurístico estocástico A, recogido por medio de la variable X, puede considerarse homogéneo en las cuatro instancias del problema del agente viajero consideradas. Dicho de otra manera, estamos interesados en estudiar si las diferencias observadas en la muestra entre las cuatro proporciones, pueden o no considerarse estadísticamente significativas.

Calculemos en primer lugar el valor del umbral c. En este caso debemos de acudir a las tablas de la distribución chi–cuadrado con 3 grados de libertad. Obtenemos que $c = 7.815$. Calculando el valor de la expresión $\sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ a partir de los datos de la tabla 3.11, obtenemos:

$$\sum_{i=1}^{f}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} =$$

$$\frac{(94-75)^2}{75}+\frac{(6-25)^2}{25}+\frac{(81-75)^2}{75}+\frac{(19-25)^2}{25}+\frac{(73-75)^2}{75}\frac{(27-25)^2}{25}+\frac{(52-75)^2}{75}\frac{(48-25)^2}{25} =$$

$$4.81+14.44+0.48+1.44+0.05+0.16+7.05+21.16 = 49.59 > 7.815 \quad (88)$$

de ahí que concluimos que la hipótesis acerca de la homogeneidad del algoritmo A en los 4 grupos debe de ser rechazada.

10.3 Test de ajuste a una distribución polinomial

La distribución chi–cuadrado se utiliza tambien como la ley de probabilidad que sigue el estadístico a partir del cual se construye el test de hipótesis para el ajuste que una distribución polinomial proporciona a una determinada muestra aleatoria.

Ejemplo 3.21
Supongamos que hemos recogido resultados de 1000 ejecuciones de un algoritmo estocástico de optimización combinatoria, que denotamos por A. Dichos resultados se relacionan con una variable aleatoria X, la cual puede tomar 5 posibles valores: 1, 2, 3, 4 y 5. El valor 1 se asignará a aquellas ejecuciones en las que el algoritmo alcanza el óptimo. El valor 2 se asignará a las ejecuciones en las que el valor alcanzado por el algoritmo A no es el óptimo, pero queda a una distancia menor del 5 por ciento del mismo. De manera análoga la variable aleatoria X tomará los valores 3, 4 y 5 cuando el valor alcanzado por el algoritmo A quede respectivamente entre el 5 y el 10 por ciento, entre el 10 y el 20 por ciento y cuando supere una distancia del 20 por ciento de dicho óptimo.

La tabla 3.13 recoge los resultados obtenidos.

Tabla 3.13: Resultados obtenidos por el algoritmo A en 1000 ejecuciones del mismo.

X	O_i
1	46
2	145
3	202
4	251
5	356

A la vista de los resultados recogidos en la tabla anterior, nos preguntamos si se puede aceptar que el modelo probabilístico generador de los datos sigue una distribución polinomial con parámetros: $p_1 = 0.05, p_2 = 0.15, p_3 = 0.20, p_4 = 0.25, p_5 = 0.35$.

Desde un punto de vista intuitivo, si fuera cierta la hipótesis que queremos testar, el número de ejecuciones que uno espera obtener para cada uno de los 5 posibles valores de la variable X son respectivamente: $E_1 = 50, E_2 = 150, E_3 = 200, E_4 = 250, E_5 = 350$.

Al igual que en los dos test anteriores (independencia y homogeneidad) cuanto más cercanos estén los valores observados en la muestra de los esperados bajo la hipótesis nula, mayor será la evidencia apoyando la hipótesis nula a contrastar. En nuestro ejemplo, los valores observados para cada uno de los 5 posibles valores de la variable X han sido: $O_1 = 46, O_2 = 145, O_3 = 202, O_4 = 251, O_5 = 356$.

Veamos a continuación la manera genérica de plantear el test de ajuste basado en la distribución chi–cuadrado, así como su aplicación al ejemplo que acabamos de introducir.

Teorema 3.13
Sea una variable aleatoria discreta, X, con f posibles valores: $x^1, \dots x^i, \dots, x^K$. Extraemos una muestra de tamaño n: $\{(x_1, \dots, x_n)\}$. Denotamos por O_i el número de casos que en la muestra toman el i-ésimo valor de X, es decir x^i, para $i = 1, \dots, K$. La región de rechazo, $R.R.$, relativa al test de hipótesis siguiente:

$$\begin{cases} H_0: & X \text{ sigue una distribución polinomial de parámetros } p_1, \dots, p_K \\ H_1: & X \text{ no sigue una distribución polinomial de parámetros } p_1, \dots, p_K \end{cases} \tag{89}$$

se define de la siguiente manera:

$$R.R. = \{(x_1, \dots, x_n) \mid \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \geq c\} \tag{90}$$

donde

- E_i denota el número de casos que bajo la hipótesis nula se espera que tomen el valor x^i
- c es un umbral que se calcula a partir de las tablas de la distribución chi–cuadrado con $K-1$ grados de libertad, a partir de la siguiente igualdad: $P(\chi^2_{K-1} > c) = \alpha$

Ejemplo 3.22
La aplicación del test anterior a los datos de la tabla 3.13 nos conduce a que $c = 9.488$ ya que $P(\chi^2_4 > 9.488) = 0.05$. Efectuando los cálculos correspondientes al estadístico obtenemos:

$$\sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} = \frac{(46-50)^2}{50} + \frac{(145-150)^2}{150} + \frac{(202-200)^2}{200} + \frac{(251-250)^2}{250} + \frac{(356-350)^2}{350} =$$

$$0.32 + 0.16 + 0.02 + 0.004 + 0.10 = 0.604 < 9.488 \tag{91}$$

Concluimos por tanto que las diferencias entre el número de casos observados en la muestra y el número de casos esperados sugeridos por el modelo de distribución polinomial teórico no son lo suficientemente grandes como para que puedan ser consideradas estadísticamente significativas. Por tanto, la evidencia recogida en la muestra no nos conduce a rechazar la hipótesis nula.

11 Test de distribución libre

En este apartado se presentarán distintos tests de distribución libre, tambien denominados tests no paramétricos. Su principal característica radica en el hecho de que pueden ser aplicados a datos provenientes de cualquier tipo de distribución probabilística.

Estos tests de distribución libre se clasifican en función del número de muestras (dos versus más de dos) y en función del tipo de muestras consideradas (apareadas versus independientes), dando origen a la tabla 3.14. En dicha tabla se muestran tambien los tests correspondientes para el caso paramétrico, en el cual se asume que los datos provienen de distribuciones normales. Nótese que el análisis de la varianza no se desarrolla en estos apuntes.

Tabla 3.14: Tabla de contingencia en un test de homogeneidad.

	tests paramétricos (distribución normal)	tests no paramétricos (distribución libre)	
		muestras aparedas	muestras independientes
2 muestras	t-test	Wilcoxon	Mann-Whitney
mas de 2 muestras	análisis de la varianza	Friedman	Kruskal-Wallis

11.1 Test para dos muestras apareadas. Test de Wilcoxon

En este punto desarrollaremos el test de Wilcoxon, el cual es de aplicación en el caso de dos muestras apareadas. Dicho test se conoce tambien como test de rangos de signos, denominación que se desprende de la información utilizada por el mismo.

TEOREMA 3.14
Sean X e Y dos variables aleatorias cualesquiera, de las que extraemos una muestra aleatoria bidimensional de tamaño n^*: $\{(x_1, y_1), \ldots, (x_{n^*}, y_{n^*})\}$. A partir de dicha muestra el siguiente test de hipótesis:

$$\begin{cases} H_0: & f_X(x) = f_Y(y) \\ H_1: & f_X(x) \neq f_Y(y) \end{cases} \tag{92}$$

se lleva a cabo por medio de los siguientes pasos:

1. Considerar $d_i = x_i - y_i$ con $(i = 1, 2, ..., n)$ siendo $n \leq n^*$ sii $d_i \neq 0$.
2. Sea $r_i =$ rango (d_i) con $i = 1, 2, ..., n$.
3. Sea $T^+ = \sum_{d_i>0} r_i$; $T^- = \sum_{d_i<0} r_i$. Se verifica: $T^+ + T^- = n(n+1)/2$. Además $E_{H_0}(T^+) = E_{H_0}(T^-) = n(n+1)/4$ y $Var_{H_0}(T^+) = Var_{H_0}(T^-) = n(n+1)(2n+1)/24$
4. Comparar T^+ con las tablas de la distribución de Wilcoxon o bien en caso de que $n > 10$ utilizar la aproximación normal

$$\frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \longrightarrow \mathcal{N}(0,1) \tag{93}$$

Es decir $R.R. = \{(x_1, y_1), \ldots, (x_{n^*}, y_{n^*}) \mid \frac{|T^+ - \frac{n(n+1)}{4}|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} > c\}$ con c verificando que: $P(|\mathcal{N}(0,1) > c) = \alpha$.

EJEMPLO 3.23
Supongamos que queramos comparar el comportamiento de dos heurísticos de optimización A y B a partir de 11 ejecuciones de los mismos. En cada una de dichas ejecuciones partimos en ambos algoritmos del mismo punto inicial, lo que nos trae como consecuencia que las dos muestras obtenidas sean apareadas. Denotamos por X e Y respectivamente a las variables aleatorias que recogen los resultados de los experimentos llevados a cabo con los heurísticos A y B. La tabla 3.15 recoge la información necesaria para llevar a cabo el test de Wilcoxon.

Tabla 3.15: Datos obtenidos en 11 ejecuciones apareadas de dos algoritmos heurísticos

ejecución	X	Y	d_i	r_i
1	94	85	9	9
2	78	65	13	10
3	89	92	-3	4
4	62	56	6	7
5	49	52	-3	4
6	78	74	4	6
7	80	79	1	1
8	82	84	-2	2
9	62	48	14	11
10	83	71	8	8
11	79	82	-3	4

Según el resultado del teorema 3.14 debemos de calcular la suma de los rangos con diferencia positiva, es decir $T^+ = \sum_{d_i>0} r_i$. El test de Wilcoxon se fundamenta en la idea de que si la hipótesis nula es cierta, el valor de T^+ no debe de distar mucho de $\frac{n(n+1)}{4}$, ya que este valor se obtendría de manera utópica (es el valor esperado) en el caso de que las dos variables, X e Y, tengan la misma función de densidad, ya que en tal caso T^+ y T^- deben de coincidir utópicamente (en sus valores esperados).

En el ejemplo se tiene: $T^+ = 9+10+7+6+1+11+8 = 52$ y $\frac{n(n+1)}{4} = \frac{11\cdot 12}{3} = 33$.

La pregunta que nos estamos haciendo es por tanto si la diferencia entre el valor de T^+ (es decir 52) y el de $\frac{n(n+1)}{4}$ (es decir 33) es lo suficientemente grande como para que no pueda considerarse que viene motivada por el azar, sino por el hecho de que las funciones de densidad de las variables X e Y son distintas.

Al comparar con $c = 1.96$ (ya que $P(|\ \mathcal{N}(0,1) > 1.96) = 0.05$), el valor $|\ \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}\ |$ obtenemos:

$$|\ \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}\ | = \frac{52-33}{\sqrt{\frac{11\cdot 12\cdot 13}{4}}} = \frac{19}{\sqrt{759}} = 0.70 < 1.96 \tag{94}$$

Concluimos por tanto que las diferencias encontradas en las muestras no llegan a ser estadísticamente significativas, de ahí que la evidencia encontrada no es lo suficientemente grande como para rechazar la hipótesis nula.

11.2 Test para dos muestras independientes. Test de Mann–Whitney

Veamos a continuación el test de Mann–Whitney, válido en el caso de que tengamos dos muestras independientes.

Teorema 3.15
Sean X e Y dos variables aleatorias cualesquiera, de las que extraemos dos muestras aleatorias independientes de tamaños respectivos n_1 y n_2. Es decir tenemos $\{x_1, \ldots, x_{n_1}\}$ e $\{y_1, \ldots, y_{n_2}\}$. Basándonos en dichas muestras, el test de hipótesis:

$$\begin{cases} H_0: & f_X(x) = f_Y(y) \\ H_1: & f_X(x) \neq f_Y(y) \end{cases} \tag{95}$$

se lleva a cabo por medio de los siguientes pasos:

1. Ordenar de menor a mayor el conjunto de las $n_1 + n_2$ observaciones.
2. Calcular R_1, R_2 sumas de los rangos de cada muestra.
3. Sean $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ y $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$. Se define $U = \max(U_1, U_2)$.
4. Si $U \geq U_s(n_1, n_2, \alpha)$ rechazamos H_0 a nivel de significación α, donde $U_s(n_1, n_2, \alpha)$ se busca en las tablas de la distribución Mann-Whitney. En caso de que $n_1 > 20$ y $n_2 > 30$ se utilizará la siguiente distribución asintótica: $Z = |U - \frac{n_1 n_2}{2}| / \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} \longrightarrow \mathcal{N}(0,1)$

Es decir $R.R. = \{(x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2})\ |\ \frac{|U - \frac{n_1 n_2}{2}|}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} > c\}$ con c verificando que: $P(|\ \mathcal{N}(0,1) > c) = \alpha$.

Ejemplo 3.24
La tabla 3.16 recoge los resultados obtenidos por dos algoritmos heurísticos de optimización A y B. Dichos resultados se reflejan como valores de dos variables aleatorias X e Y.

Tabla 3.16: Resultados obtenidos por ejecuciones independientes de dos algoritmos A y B.

ejecución	X	r	r	Y	ejecución
1	78	7	9	110	1
2	64	4	5	70	2
3	75	6	3	53	3
4	45	1	2	51	4
5	82	8			

Vamos a aplicar el resultado asintótico presentado en el teorema anterior, a pesar de que los tamaños de las muestras no verifican las condiciones de aplicabilidad. Obtenemos $R_1 = 7+4+6+1+8 = 26$, $R_2 = 0+5+3+2 = 19$, $U_1 = 5\cdot 4 + \frac{5\cdot 6}{2} - 26 = 20+15-26 = 9$, $U_2 = 5\cdot 4 + \frac{4\cdot 5}{2} - 19 = 20+10-19 = 11$, y $U = \max(9,11) = 11$. El valor del estadístico será:

$$| \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} | = \frac{11 - \frac{5\cdot 4}{2}}{\sqrt{\frac{5\cdot 4\cdot (5+4+1)}{12}}} = \frac{1}{\sqrt{16.66}} < 1.96 \tag{96}$$

Por lo tanto concluimos que la evidencia obtenida en la muestra no es lo suficientemente concluyente como para rechazar la hipótesis nula.

11.3 Test para más de dos muestras aparedas. Test de Friedman

En este punto se mostrará el test de Friedman, el cual se aplica en el caso de plantearnos por un contraste en cuya hipótesis nula se afirme la igualdad de más de dos funciones de densidad, a partir de más de dos muestras aparedas.

Teorema 3.16
Sean $X_1, \ldots, X_K$ variables aleatorias cualesquiera. A partir de una muestra aleatoria apareada de tamaño n obtenida de la variable aleatoria K dimensional $(X_1, \ldots, X_K)$ –véase tabla 3.17– planteamos el siguiente test de hipótesis:

$$\begin{cases} H_0: & f_{X_1}(x_1) = f_{X_2}(x_2) = \ldots = f_{X_K}(x_K) \\ H_1: & \text{Las } K \text{ funciones de densidad no son iguales} \end{cases} \tag{97}$$

Tabla 3.17: Muestra apareada de tamaño n obtenida de la variable aleatoria k dimensional $(X_1, \ldots, X_K)$

X_1	X_2	$\ldots$	X_j	$\ldots$	X_K
x_{11}	x_{21}	$\ldots$	x_{j1}	$\ldots$	x_{K1}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_{1l}	x_{2l}	$\ldots$	x_{jl}	$\ldots$	x_{Kl}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_{1n}	x_{2n}	$\ldots$	x_{jn}	$\ldots$	x_{Kn}

Dicho test de hipótesis se resuelve a partir de los siguientes pasos:

1. Para cada fila asignar un rango a cada observación comenzando por 1 y terminando por K. Es decir $r_{jl} = rango(x_{jl})$
2. Sumar todos los rangos para cada variable. Sea R_j la suma de los rangos de la j-ésima variable. $R_j = \sum_{l=1}^{n} r_{jl}$. Bajo H_0, los R_j deben de variar poco.
3. Bajo H_0 el estadístico $S = \frac{12}{nK(K+1)}(\sum_{l=1}^{K} R_j^2) - 3n(K+1)$ se distribuye aproximadamente como una Chi-cuadrado con $K-1$ grados
4. Región de rechazo $S > \mathcal{X}^2_{K-1;\alpha}$

Ejemplo 3.25
Supongamos 4 algoritmos heurísticos A, B, C, y D cada uno de los cuales es ejecutado 10 veces. En cada una de estas 10 ejecuciones, cada uno de los 4 heurísticos parte de la misma inicialización, de ahí que los resultados puedan considerarse como correspondientes a muestras apareadas. La tabla 3.18 recoge dichos resultados así como los valores de los rangos necesarios para llevar a cabo el test de Friedman.

Tabla 3.18: Resultados obtenidos con 4 algoritmos heurísticos en una muestra apareada de tamaño 10.

	X_1	r_1	X_2	r_2	X_3	r_3	X_4	r_4
1	8.5	3	8.6	4	8.2	1	8.4	2
2	9.8	4	9.7	3	9.4	1	9.6	2
3	7.9	2	8.1	3	7.5	1	8.2	4
4	9.7	3	9.8	4	9.6	1.5	9.6	1.5
5	6.2	1	6.8	3	6.9	4	6.5	2
6	8.9	3	9.2	4	8.1	1	8.7	2
7	9.2	3.5	9.2	3.5	8.7	1	8.9	2
8	8.4	1.5	8.5	3	8.4	1.5	8.6	4
9	9.2	2	9.6	4	8.9	1	9.5	3
10	8.8	2	9.2	3	8.6	1	9.3	4

A partir de los resultados reflejados en la tabla 3.18, obtenemos que: $R_1 = 25$, $R_2 = 34.5$, $R_3 = 14$, $R_4 = 26.5$. El valor del estadístico $S = \frac{12}{nK(K+1)}(\sum_{l=1}^{K} R_j^2) - 3n(K+1)$ resulta ser: $S = \frac{12}{10 \cdot 4 \cdot 5}(25^2 + 34.5^2 + 14^2 + 26.5^2) - 3 \cdot 10 \cdot 5 = 12.81$. Al comparar dicho valor con el obtenido en las tablas de la distribución chi-cuadrado con 3 grados de libertad, y con $\alpha = 0.01$, es decir con 11.32, ya que $(P(\chi_3^2 > 11.32) = 0.01)$ obtenemos que $12.81 > 11.32$, y por tanto las diferencias entre las muestras correspondientes a los cuatro heurísticos han resultado ser estadísticamente significativas, de ahí que se rechaze la hipótesis nula de igualdad entre las 4 funciones de densidad.

11.4 Test para más de dos muestras independientes. Test de Kruskal–Wallis

En este punto vamos a desarrollar el test de Kruskal–Wallis válido para el caso de contrastar la igualdad de más de dos funciones de densidad en el caso de muestras independientes.

Teorema 3.17
Sean $X_1, \ldots, X_K$ variables aleatorias cualesquiera. A partir de k muestras independientes de tamaños respectivos $n_1, n_2, \ldots n_K$ con $N = \sum_{i=1}^{K} n_i$ –véase tabla 3.19– planteamos el siguiente test de hipótesis:

$$\begin{cases} H_0: & f_{X_1}(x_1) = f_{X_2}(x_2) = \ldots = f_{X_K}(x_K) \\ H_1: & \text{Las } K \text{ funciones de densidad no son iguales} \end{cases} \tag{98}$$

Tabla 3.19: K muestras independientes de tamaños respectivos $n_1, n_2, \ldots n_K$ extraidas de las variables aleatorias $X_1, \ldots, X_K$.

X_1	X_2	$\ldots$	X_j	$\ldots$	X_K
x_{11}	x_{21}	$\ldots$	x_{j1}	$\ldots$	x_{K1}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\ldots$	$x_{2_{n_2}}$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$x_{1_{n_1}}$			$\ldots$	$\ldots$	$x_{K_{n_K}}$
			x_{jn_j}		

Dicho test se resuelve a partir de los siguientes pasos:

1. Ordenar de menor a mayor el conjunto de las $\sum_{i=1}^{K} n_i = N$ observaciones.

2. Calcular $R_i = \sum_{w=1}^{n_i} r(x_{iw})$ con $i = 1, ..., K$.

3. Estudiar la disparidad de las R_i a partir del estadístico:

   $H = \frac{12}{N(N+1)}(\sum_{j=1}^{K} \frac{(R_j)^2}{n_j}) - 3(N+1)$, el cual se distribuye aproximadamente como una Chi-cuadrado con $K-1$ grados

4. Región de rechazo a nivel α: $H > \mathcal{X}^2_{K-1;\alpha}$

Ejemplo 3.26
Supongamos 4 heurísticos de optimización A, B, C y D, los cuales son ejecutados $n_1 = 5, n_2 = 6, n_3 = 6$ y $n_4 = 5$ veces respectivamente. Los valores obtenidos se recogen en la tabla 3.20, en la cual se indica tambien el valor del rango correspondiente.

Tabla 3.20: Resultados obtenidos con 4 algoritmos heurísticos en muestras independientes de tamaños respectivos $n_1 = 5, n_2 = 6, n_3 = 6$ y $n_4 = 5$.

X_1	r_1	X_2	r_2	X_3	r_3	X_4	r_4
1.19	15	1.08	4.5	0.98	2	1.12	7.5
1.05	3	1.23	17.5	1.19	15	1.14	10
1.14	10	1.26	20	1.08	4.5	1.31	22
1.25	19	1.10	6	0.93	1	1.12	7.5
1.29	21	1.18	12.5	1.23	17.5	1.19	15
		1.14	10	1.18	12.5		

A partir de los resultados de la tabla 3.20 obtenemos: $R_1 = 68, R_2 = 70.5, R_3 = 52.5, R_4 = 62$. Dichos valores nos sirven para calcular el valor del estadístico $H = \frac{12}{22 \cdot 23}(\frac{68^2}{5} + \frac{70.5^2}{6} + \frac{52.5^2}{6} + \frac{62^2}{5}) - 3 \cdot 23 = 1.70$. Al comparar dicho valor con el percentil ($\alpha = 0.05$) de las tablas de la chi–cuadrado con 3 grados de libertad, es decir $c = 7.82$, obtenemos que $1.70 < 7.82$, de ahí que concluyamos que las diferencias encontradas en las 4 muestras independientes no son estadísticamente significativas y decidamos no rechazar la hipótesis nula de igualdad de funciones de densidad.

Ejercicios

1. Con objeto de testar si la presión sanguinea de los individuos que padecen una determinada enfermedad puede ser considerada igual que la de individuos libres de dicha enfermedad, se han escogido al azar 12 individuos con dicha enfermedad, y 10 de entre los que no la padecen, obteniéndose los resultados expresados en la Tabla 1. Establecer si la muestra pertenece a la región de aceptación o a la región de rechazo. Construir el test a nivel $\alpha = 0.05$.

2. A partir de una muestra aleatoria de tamaño n, obtener la región de rechazo óptima a nivel α, para los siguientes test de hipótesis, con hipótesis nula y alternativa simples:

   - Siendo λ el parámetro de una distribución de Poisson

     $$H_0 : \lambda = \lambda_0$$

     $$H_1 : \lambda = \lambda_1$$

   - Siendo p el parámetro de una distribución $B(m, p)$

     $$H_0 : p = p_0$$

     $$H_1 : p = p_1$$

muestra	variable(s)
1	167
1	167
1	156
1	188
1	178
1	164
1	175
1	145
1	182
1	190
1	156
1	177
2	140
2	170
2	153
2	151
2	167
2	166
2	133
2	155
2	164
2	144

Tabla 1: Presiones sanguineas de individuos enfermos y sanos

- Siendo p el parámetro de una distribución de Bernouilli

$$H_0 : p = p_0$$

$$H_1 : p = p_1$$

- Siendo μ el parámetro de una distribución normal con varianza conocida σ^2

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu = \mu_1$$

- Siendo Θ el parámetro de una distribución exponencial

$$H_0 : \Theta = \Theta_0$$

$$H_1 : \Theta = \Theta_1$$

3. A partir de una muestra de tamaño 49 proveniente de una distribución normal con parámetro μ desconocido y parámetro σ =3, se ha obtenido una media aritmética de $\overline{X}$ =14,7.
Testear la hipótesis nula $H_0 : \mu = 10$ frente a la alternativa $H_1 : \mu = 16$, con un nivel de significación de $\alpha = 0,05$

4. En la tabla adjunta, se puede observar la distribución del número de suspensos obtenidos en la calificación de Junio por los 300 alumnos cursando el primer curso de Informática.

Número de suspensos	0	1	2	3	4
Número de alumnos	50	100	80	50	20

Testear la hipótesis nula de que la variable aleatoria que mide el número de suspensos obtenidos por un alumno de primer curso en la convocatoria de Junio, se distribuye según un modelo binomial.
Plantear el problema utilizando dos test distintos.

5. Se sabe que una bolsa contiene una bola roja y cuatro blancas o alternativamente, cuatro rojas y una blanca. Se extrae una bola. La hipótesis de que una bola es roja y cuatro son blancas no se rechaza si y sólo si la bola extraida es blanca.
   (i) Estudio e interpretación de los dos tipos de errores, asi como del concepto de potencia del test.
   (ii) Supóngase que la alternativa es tres bolas rojas y dos blancas. Rehacer el estudio.

6. La Tabla 2 recoge información relativa a un test de conocimientos efectuado sobre un conjunto de 10 individuos, antes (variable X) y después (variable Y) de haber seguido un cursillo. Podemos afirmar que el cursillo ha sido efectivo?. Se entiende que la puntuación es directamente proporcional al conocimiento.

enfermo	X (antes)	Y (después)
1	149	152
2	78	94
3	80	99
4	94	85
5	78	165
6	84	92
7	60	89
8	82	84
9	62	48
10	83	91

Tabla 2: Resultados antes y después de seguir un cursillo

7. Construir la región de rechazo intuitiva para los siguientes test de hipótesis relacionados con una variable aleatoria X siguiendo una distribución normal con esperanza matemática μ desconocida y varianza σ^2 conocida. El test se construye a nivel de significación α, a partir de una muestra aleatoria de tamaño n.

   (a)

   $$H_0 : \mu = \mu_0$$

   $$H_1 : \mu \neq \mu_0$$

   (b)

   $$H_0 : \mu = \mu_0$$

   $$H_1 : \mu > \mu_0$$

   (c)

   $$H_0 : \mu = \mu_0$$

   $$H_1 : \mu < \mu_0$$

   (d)

   $$H_0 : \mu < \mu_0$$

   $$H_1 : \mu \geq \mu_0$$

   (e)

   $$H_0 : \mu > \mu_0$$

   $$H_1 : \mu \leq \mu_0$$

8. Construir la región de rechazo intuitiva para los siguientes test de hipótesis relacionados con una variable aleatoria X siguiendo una distribución binomial $B(m, p)$, siendo p un parámetro desconocido. El test se construye a nivel de significación α, a partir de una muestra aleatoria de tamaño n.

   (a)

   $$H_0 : p = p_0$$

   $$H_1 : p \neq p_0$$

   (b)

   $$H_0 : p = p_0$$

   $$H_1 : p > p_0$$

   (c)

   $$H_0 : p = p_0$$

   $$H_1 : p < p_0$$

   (d)

   $$H_0 : p < p_0$$

   $$H_1 : p \geq p_0$$

   (e)

   $$H_0 : p > p_0$$

   $$H_1 : p \leq p_0$$

9. Construir la región de rechazo intuitiva para los siguientes test de hipótesis relacionados con una variable aleatoria X siguiendo una distribución de Poisson con parámetro λ desconocido. El test se construye a nivel de significación α, a partir de una muestra aleatoria de tamaño n.

   (a)

   $$H_0 : \lambda = \lambda_0$$

   $$H_1 : \lambda \neq \lambda_0$$

   (b)

   $$H_0 : \lambda = \lambda_0$$

   $$H_1 : \lambda > \lambda_0$$

   (c)

   $$H_0 : \lambda = \lambda_0$$

   $$H_1 : \lambda < \lambda_0$$

   (d)

   $$H_0 : \lambda < \lambda_0$$

   $$H_1 : \lambda \geq \lambda_0$$

   (e)

   $$H_0 : \lambda > \lambda_0$$

   $$H_1 : \lambda \leq \lambda_0$$

10. Una urna contiene 10 bolas y se desea probar la hipótesis de que 2 bolas son rojas y 8 blancas, contra la alternativa de que más de 2 son rojas. Se extraen 2 bolas sin reemplazamiento y se rechaza la hipótesis, si y sólo si, las 2 bolas extraidas son rojas.
(i) Estudiar la probabilidad de cometer un error de tipo I, asi como la función potencia del test.
(ii) Rehacer el ejercicio para el caso en que las extracciones se efectúen con reemplazamiento.

11. El resultado de uno de los famosos experimentos de Mendel en cruzamiento de plantas fué de 355 guisantes amarillos y 123 verdes. Probar si este resultado está de acuerdo con la teoría de Mendel, de acuerdo a la cual el resultado debe ser de $3:1$.

12. Teniendo en cuenta la Tabla 3, probar que la distribución subyacente a la variable aleatoria que mide el número de muertes provocadas por el coceo de un caballo en 10 Cuerpos del Ejército Prusiano durante el periodo 1875-1894, sigue una distribución de Poisson.

X	Observados
1	167
0	109
1	65
2	22
3	3
4	1
5	0

Tabla 3: Número de casos observados con x muertes en cada cuerpo.

13. Con el fin de comparar dos métodos que sirven para determinar el contenido de almidón en las patatas, se cortan en dos mitades 16 patatas, y a cada una de las dos mitades se les aplica uno de los métodos. Las diferencias fueron:
2, 0, 0, 1, 2, 2, 3, -3, 1, 2, 3, 0, -1, 1, 1, -2.
Probar la hipótesis de que la diferencia entre los métodos no es significativa.

14. Teniendo en cuenta la Tabla 4, estudiar la independencia entre la estatura de los bebés al nacer - variable X - y la circunferencia de la cabeza de los mismos - variable Y -.

		X		
		47-49	50-52	53-55
Y	32-35	40	36	2
	36-39	0	14	7

Tabla 4: Estatura y circunferencia de los bebés al nacer.

15. Los resultados obtenidos por cuatro algoritmos en 10 problemas distintos aparecen expresados en la Tabla 5. Se puede afirmar que los 4 cuatro algoritmos tienen comportamientos iguales, o las diferencias obtenidas no pueden ser atribuibles al azar?

16. Los resultados de 4 tratamientos (A, B, C, D) aplicados a enfermos independientes se muestran en la Tabla 6. Testear la hipótesis de igualdad de tratamiento a nivel de significación $\alpha = 0.05$.

	Algoritmo			
	1	2	3	4
1	18.5	28.6	18.2	28.4
2	19.8	29.7	19.4	29.6
3	17.9	28.1	27.5	28.2
4	19.7	19.8	29.6	29.6
5	16.2	16.8	26.9	26.5
6	18.9	29.2	28.1	28.7
7	19.2	19.2	28.7	18.9
8	18.4	28.5	18.4	18.6
9	19.2	19.6	18.9	19.5
10	8.8	29.2	18.6	19.3

Tabla 5: Resultados obtenidos por cuatro algoritmos en 10 problemas distintos.

Tratamiento			
A	B	C	D
7	8	7	9
10	12	10	9
11	10	10	9
14	14	11	10
10	11	12	11
	11	13	10
	13	12	16
	12	13	
	13	11	
	12		
	13		

Tabla 6: Resultados obtenidos por cuatro tratamientos en pacientes no " apareados ".